SEXUAL SELECTION AND SELECTIVE CONSTRAINT:

THE EVOLUTION OF A NOVEL REPRODUCTIVE GLAND

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Findley Ransler Finseth

August 2013

SEXUAL SELECTION AND SELECTIVE CONSTRAINT: THE EVOLUTION OF A NOVEL REPRODUCTIVE GLAND

Findley Ransler Finseth, Ph. D.

Cornell University 2013

Sexual selection acts on an organism's ability to secure mates. When females mate multiply, it sets the stage for sexual selection to continue after mating. This post-mating sexual selection arises from variation in individual's ability to obtain fertilizations and shapes both quantitative and qualitative variation in traits. Newly arisen traits that are qualitatively different from previous ones are considered "novel". This dissertation focuses on the evolution of novel traits involved in post-mating sexual selection. Male Japanese quail (*Coturnix japonica*) possess a foam gland that secretes a viscous fluid they whip into an airy, meringue-like foam. When males mate with females, they transfer semen along with a large quantity of foam. Both the foam gland and the foamy secretion are novel traits. We characterize *i*) the current function of foam in post-mating sexual selection, *ii*) the molecular basis of the foam, and *iii*) the selective pressures that shaped the foam gland and foam gland genes. Chapter 1 finds that foam influences the outcome of sperm competition (a form of post-mating sexual selection) by benefitting the sperm of males who transfer both sperm and foam. Chapter 2 investigates the utility of two leading next-generation sequencing technologies for transcriptome characterization and gene expression quantification and lays the foundation for the genomic work in Chapters 3 and 4. Chapter 3 compares the selective pressures shaping genes with enriched expression in the foam

gland, testis, or liver or those that are ubiquitously expressed. We find that foam gland genes reveal surprising levels of selective conservation. Chapter 4 characterizes the molecular basis of the foam proteome. This chapter finds that foam genes were largely co-opted from genes with ancient evolutionary origins followed by sequence conservation, but some individual genes evolved rapidly. Ultimately, novel reproductive traits involved in post-mating sexual selection can emerge from re-purposing of conserved genes followed by selective constraint.

BIOGRAPHICAL SKETCH

Findley Finseth (née Ransler) was raised in Paducah, KY and first became interested in science at Paducah Tilghman High School. There, Nelta Freeman (AKA 'Nelta Beast') taught a challenging and forward-looking biology course that incorporated gel electrophoresis, polymerase chain reaction, and, famously, cat dissections. It motivated her to continue studying science at the University of Virginia (UVA), where she majored in biology. Highlights at UVA included excellent classes, a hands-on marine ecology field course in the Bahamas (thanks, Dr. Diehl!), and a summer spent studying the behavior of marmots (*i.e.*, 'marmoteering') at the Rocky Mountain Biological Laboratory under Dr. Dan Blumstein. During her senior year, she worked as a lab technician in Dr. Janis Antonovics' plant disease ecology lab where she pursued an independent project investigating the mitochondrial inheritance of anther smut. It was in this position under the excellent guidance and steady encouragement of Dr. Michael Hood where Findley began seriously considering a career in science. After college, Findley worked briefly with Julie Lease in Hawaii at the Kilauea Field Station in Hawaii Volcanoes National Park on avian malaria before joining the Rocky Mountain Center for Conservation Genetics and Systematics (RMCCGS) in Denver, Colorado. At RMCCGS, she applied population genetic techniques towards conservation efforts of endangered species, chiefly the trumpeter swan, under Drs. Sara Oyler-McCance and Tom Quinn. Here, Findley solidified her plans to attend graduate school to study evolutionary genetics.

She met Dr. Rick Harrison during the opening SNEEB of the 2006 recruitment weekend for Cornell's Ecology and Evolutionary Biology department. They talked at length

about the evolution of reproductive proteins in crickets over a couple of glasses of beer (Findley) or wine (Rick). That conversation inspired Findley to pursue a rotation with Rick the summer after her first year in graduate school, which ultimately let to many happy days spent studying the evolution of reproductive proteins found in foam of male Japanese quail under Rick's guidance. Upon finishing her degree at Cornell University, Findley will move to Missoula, Montana to pursue a post-doctoral position with Dr. Lila Fishman, where she will study evolutionary genetics and reproductive biology in plants. Finally, Findley feels damn lucky that her scientific career has been peppered with many inspiring role models who were both brilliant scientists and first-rate mentors.

To Beatrix Finseth,

for bringing biology home.

ACKNOWLEDGMENTS

In my opinion, one of the main factors dictating the quality of a graduate experience is your advisor. I could not have had a better one. Rick Harrison cares deeply about his students both personally and professionally and it has left a major impact on me. He encourages us to make decisions that ensure that our work contributes to a full, rewarding life. I appreciate Rick's perceptiveness, careful attention to detail, generosity with financial support, facilitation of academic independence (he let me work on a bird!) and the healthy dose of skepticism he brings to all matters scientific. Rick sets a great example of balancing work with all the good things in life and possesses a rare combination of emotional and intellectual intelligence. I am grateful that he was my advisor.

Elizabeth Adkins-Regan provided invaluable guidance and encouragement throughout my dissertation. She initially introduced me to Japanese quail/foam and offered generous financial support and unflagging enthusiasm for my work throughout my dissertation. Her ability to synthesize across many disciplines of organismal biology and ask 'big questions' is an inspiration. Brian Lazzaro laid the foundation for my understanding of a quantitative approach to evolutionary biology. He has been an excellent teacher and I am always wowed by the clarity he brings to the fuzziness of biology.

This dissertation would not have been possible without the logistical help of many people. In particular, Eliana Bondra and Stephanie Iacovelli have been true colleagues on many aspects of this research. Steve Bogdanowicz has provided a steady stream of answers to all lab-related questions. The Harrison lab has been a source of thought-provoking discussions, critical feedback, and friendship, with particular thanks to Luana

Maroja, Marie Nydam, Petra Deane, Erica Larson, Ben Hamilton, Ryan Thum, Jamie Walters, Steve Bogdanowicz, Jose Andres, and Eliana Bondra. Dave Cerasale and Nicole Baran spent many hours handling birds with care and expertise. The Cornell Institute of Biotechnology was integral to the success of this project, especially Jennifer Mosher, Peter Schweitzer, Shang Zheng, James McCardle, Qi Sun, and Robert Bukowski. Tim van Deusen, Linda Vann, Percy Smith, and Stephanie Van Martin provided top notch care of quail. Rich Meisel, Geoff Findlay, Brooks Miner, Morgan Mouchka, Nancy Chen, Jamie Walters, Chris Dalton, Mike Booth, Ezra Lencer, Erica Larson, Jeremy Searle, and Dave Cerasale gave me excellent feedback and/or advice.

When I visited Cornell's EEB department for recruitment weekend, I was fully intending on going to a different university. Upon meeting the students in EEB, my plans changed. Seven years later, my fellow graduate students are still the best part of this gig. Thanks in particular to Erica Larson, Morgan Mouchka, Chris Dalton, Ezra Lencer, Dave Cerasale, Mike Booth, Paulo Llambias, Sarah States, Ben Hunt, Danica Lombardozzi, Rayna Bell, and Ginny Howick. You made Ithaca feel like home.

I would also like to thank my parents who always believed I could do anything. They brought me up in an environment of unconditional love and support, providing me with the confidence to take on challenging tasks. Abby, Anna, and Chip have traveled many long hours to spend cherished time in Ithaca and have provided a network of encouragement and love throughout my life.

Finally, my greatest thanks goes to Ryan and Beatrix. You have been my tireless support system, my source of motivation, and my daily reminders about how wonderful life can be.

TABLE OF CONTENTS

BIOGRAPHICAL SKETCH ..... **III**

ACKNOWLEDGMENTS ..... **V**

LIST OF FIGURES ..... **VII**

LIST OF TABLES ..... **VIII**

CHAPTER 1 ..... **1**

A NON-SEMEN COPULATORY FLUID INFLUENCES THE OUTCOME OF SPERM COMPETITION *1*

Abstract *2*

Introduction *2*

Methods *9*

Results *19*

Discussion *23*

Acknowledgements *33*

References *34*

CHAPTER 2 ..... **41**

A COMPARISON OF NEXT-GENERATION SEQUENCING APPROACHES FOR TRANSCRIPTOME ASSEMBLY AND UTILITY FOR RNA-SEQ IN A NON-MODEL SPECIES *41*

Abstract *42*

Introduction *43*

Methods *46*

Results and Discussion *54*

Conclusions *70*

Acknowledgements *72*

References *73*

CHAPTER 3 ..... **79**

PHENOTYPIC DIVERGENCE AROSE WITHOUT MAJOR PROTEIN-CODING SEQUENCE DIVERGENCE IN A NOVEL REPRODUCTIVE GLAND *79*

Abstract *80*

Introduction *80*

Methods *85*

Results *93*

Discussion *100*

Conclusions *113*

Acknowledgements *114*

References *115*

CHAPTER 4 ..... **125**

COMBINED PROTEOMICS AND RNA-SEQ REVEAL CONTRASTING PATTERNS OF CONSTRAINT AND DIVERGENCE DURING THE EVOLUTION OF A NOVEL REPRODUCTIVE FLUID *125*
Abstract *126*
Introduction *126*
Methods *131*
Results *141*
Discussion *149*
Conclusion *163*
Acknowledgements *164*
References *165*

APPENDIX 174

SUPPLEMENTARY INFORMATION *174*
Chapter 1 ..... *175*
Chapter 2 ..... *178*
Chapter 3 ..... *179*
Chapter 4 ..... *190*

LIST OF FIGURES

CHAPTER 1 ..... **1**
FIGURE 1.1: Foam gland system of male Japanese quail *7*
FIGURE 1.2: Design for sperm competition experiment *10*
FIGURE 1.3: Paternity of offspring from competitively mated males after foam manipulation *22*
FIGURE 1.4: Proportion of trials with any fertilizations after foam manipulation *25*

CHAPTER 2 ..... **41**
FIGURE 2.1: Contig length distributions from various transcriptomes *58*
FIGURE 2.2: Venn diagram of the number of orthologs from various transcriptomes *62*
FIGURE 2.3: Ortholog hit ratios from various transcriptomes *64*
FIGURE 2.4: Ortholog hit ratios versus ortholog length *66*
FIGURE 2.5: RNA-Seq mapping assessment *68*

CHAPTER 3 ..... **79**
FIGURE 3.1: Venn diagram of tissue-specific genes *95*
FIGURE 3.2: Partial correlation of expression, enrichment, and evolutionary rate *99*
FIGURE 3.3: Significant shifts in selective pressure along the quail or chicken lineage *105*
FIGURE 3.4: Magnitude of shifts in selective pressure along the quail or chicken lineage *106*

CHAPTER 4 ..... **125**
FIGURE 4.1: Foam glands and foam from three photoperiod and hormone treatments *130*
FIGURE 4.2: Foam gland response to photoperiod and hormone treatments *142*
FIGURE 4.3: Correlation of foam protein abundance and gene expression level *144*
FIGURE 4.4: Proportion of transcript categorizations represented by orthologs *146*
FIGURE 4.5: Evolutionary rates of foam proteins *158*
FIGURE 4.6: Evolutionary rates of secreted foam proteins *159*
FIGURE 4.7: Evolutionary origins of foam proteins *160*

APPENDIX **174**
FIGURE S3.1: Ortholog hit ratios from filtered transcriptome *186*
FIGURE S3.2: Normalized numbers of reads for different tissues *187*
FIGURE S3.3: Multidimensional scaling plot of RNA-Seq samples *188*
FIGURE S3.4: Residuals from models regressing dN on dS across tissues *189*

LIST OF TABLES

CHAPTER 1 ..... **1**
TABLE 1.1: Model selection for sperm competition experiment *21*

CHAPTER 2 ..... **41**
TABLE 2.1: Standard metrics of transcriptome quality *56*
TABLE 2.2: Statistics about contigs with or without open reading frames *56*
TABLE 2.3: Errors in transcriptome assembly *56*

CHAPTER 3 ..... **79**
TABLE 3.1: Patterns of protein evolution derived from quail:chicken orthologs *97*
TABLE 3.2: Genes under positive selection and with accelerated evolution in quail *101*

CHAPTER 4 ..... **5**
TABLE 4.1: Twenty most abundant foam proteins *147*
TABLE 4.2: Enrichment of clustered biological processes and molecular functions in foam *148*

APPENDIX **174**
TABLE S1.1: Details about primers for paternity analyses *175*
TABLE S1.2: Parameter estimates from best models *176*
TABLE S1.3: Overall and group-specific paternity means and/or proportions *177*
TABLE S2.1: Summary statistics of raw data *178*
TABLE S3.1: Sequencing reads used for RNA-Seq analyses *183*
TABLE S3.2: Standard metrics of filtred transcriptome quality *184*
TABLE S3.3: RT-qPCR validation of RNA-Seq data *185*
TABLE S4.1: RT-qPCR validation of RNA-Seq from foam glands *193*
TABLE S4.2: Foam production response to hormone and photoperiod treatment *194*
TABLE S4.3: Annotation of foam proteins *195*
TABLE S4.4: P-values from ortholog comparisons among FG expression classes *252*
TABLE S4.5: P-values from ortholog comparisons among tissues *253*
TABLE S4.6: P-values from evolutionary origin comparisons among FG expression classes *254*
TABLE S4.3: P-values from evolutionary origin comparisons among tissues *255*

CHAPTER 1

A NON-SEMEN COPULATORY FLUID INFLUENCES THE OUTCOME OF SPERM COMPETITION

Published as: Findley R. Finseth, Stephanie R. Iacovelli, Richard G. Harrison, and Elizabeth K. Adkins-Regan 2013. A non-semen copulatory fluid influences the outcome of sperm competition. Journal of Evolutionary Biology *in press.*

Abstract

Sperm competition is a powerful and widespread evolutionary force that drives the divergence of behavioral, physiological, and morphological traits. Elucidating the mechanisms governing differential fertilization success is a fundamental question of sperm competition. Both sperm and non-sperm ejaculate components can influence sperm competition outcomes. Here, we investigate the role of a non-semen copulatory fluid in sperm competition. Male Japanese quail possess a gland that makes meringue-like foam. Males produce and store foam independent of sperm and seminal fluid, yet transfer foam to females during copulation. We tested whether foam influenced the outcome of sperm competition by varying foam state and mating order in competitive matings. We found that the presence of foam from one male decreased the relative fertilization success of a rival, and that foam from a given male increased the probability he obtained any fertilizations. Mating order also affected competitive success. Males mated first fertilized proportionally more eggs in a clutch and had more matings with any fertilizations than subsequent males. We conclude that the function of foam in sperm competition is mediated through the positive interaction of foam with a male's sperm, and we speculate whether the benefit is achieved through improving sperm storage, fertilizing efficiency, or retention. Our results suggest males can evolve complex strategies to gain fertilizations at the expense of rivals as foam, a copulatory fluid not required for fertilization, nevertheless has important effects on reproductive performance under competition.

Introduction

Identifying the causes of variation in male reproductive success is central to understanding the process of sexual selection. Historically, work on sexual selection

focused on traits used for acquiring mates (Andersson, 1994; Wang *et al.*, 2009; Birkhead, 2010). Following the publication of Parker's seminal paper exploring insects' diverse adaptations to sperm competition (when ejaculates from multiple males overlap in a female's reproductive tract and compete for fertilization for a given set of ova), research has emphasized traits conferring an advantage in sexual selection that occurs after mating (post-copulatory sexual selection) (Parker, 1998; Birkhead & Møller, 1998; Simmons, 2001; Birkhead & Pizzari, 2002; Snook, 2005; Pizzari & Parker, 2008; Birkhead, 2010). It is clear from this body of work that sperm traits evolve to increase a male's fertilization success at the expense of his competitors. Both theoretical and empirical research also demonstrate that sperm phenotype, together with number, production, and allocation strategies of sperm, are targets of sperm competition (Parker, 1998; Snook, 2005; Pizzari & Parker, 2008).

While attributes of sperm clearly impact the outcome of sperm competition, there is increasing evidence that non-sperm ejaculate components also play a key role in post-copulatory sexual selection (Ramm *et al.*, 2005; Cameron *et al.*, 2007; Wigby *et al.*, 2009; Perry & Rowe, 2010; Fedorka *et al.*, 2010; Parker & Pizzari, 2010; Sirot *et al.*, 2011; Lemaitre *et al.*, 2011; Simmons & Fitzpatrick, 2012). One important class of non-sperm ejaculate traits affecting competitive fertilization success are seminal fluid proteins (Wolfner, 1997; Chapman, 2001; Simmons, 2001; Chapman & Davies, 2004; Poiani, 2006). Males of internally fertilizing species transfer sperm to females along with a cocktail of proteins comprising the seminal fluid. Seminal fluid proteins affect sperm competition outcomes in several ways, including modifying female receptivity and behavior, forming copulatory plugs, mediating sperm transport and storage, preventing displacement from a

subsequent male's sperm, and contributing to sperm motility (Wolfner, 1997; Poiani, 2006; Chapman, 2001; Simmons, 2001; Chapman & Davies, 2004; Pitnick *et al.*, 2008, Dean *et al.* 2011). Theoretical models predict that strategic allocation of seminal fluid proteins can enhance the relative fitness of competing males, for example by exploiting a protein with a beneficial function (*e.g.*, fecundity-stimulating or buffering of the hostile reproductive tract) transferred by a rival male (Hodgson & Hosken, 2006; Cameron *et al.*, 2007; Alonzo & Pizzari, 2010).

Recent empirical work highlights the remarkable sophistication of evolutionary responses of non-sperm components of the ejaculate to sperm competition. In *Drosophila melanogaster*, males selected for larger accessory glands (a major site of seminal fluid production) produce more of a fecundity-stimulating seminal fluid protein (ovulin) and are more successful in competitive mating scenarios (Wigby *et al.*, 2009). Further, *D. melanogaster* males not only strategically adjust the volume of seminal fluid they transfer to females according to the intensity of sperm competition, but they adaptively tailor the amount of specific proteins in their ejaculate according to a female's mating status (Wigby *et al.*, 2009; Sirot *et al.*, 2011). Comparative evidence across rodents suggests that larger copulatory plugs, made by larger accessory glands which produce the coagulating fluids, may be adaptations to sperm competition in species with more promiscuous females (Ramm *et al.*, 2005). In the bank vole, an experimental rise in sperm competition intensity increases the size of a major male accessory gland (the seminal vesicle), but not testis size, sperm production, or sperm motility (Lemaitre *et al.*, 2011). In birds, female zebra finch mated to a vasectomized male, therefore receiving seminal fluid but no sperm, are no more likely to seek extra-pair copulations than those mated to intact males who transfer sperm

(Birkhead & Fletcher, 1995). Fish (Smith & Ryan, 2011), fowl (Pizzari *et al*., 2007), and humans (Kilgallon & Simmons, 2005) produce ejaculates with higher quality sperm or a larger proportion of motile sperm in response to increases in the number of potential rivals or variation in the social status of competing males, which are correlates for perceived risk of sperm competition. The most promising candidate for a mechanism underlying this phenomenon is that males adjust their seminal fluid to increase sperm swimming speed, as occur when dominant male fowl mate with high-quality females (Cornwallis & O'Connor, 2009). Further, seminal fluid can differentially affect the sperm performance and fertilizing efficiency of rivals according to a male's reproductive tactic (territorial or sneaker), which determines his risk of sperm competition (Locatello *et al.,* 2013).

Although non-sperm ejaculate components are taxonomically widespread and clearly affect sperm competition dynamics, their role is largely understudied when compared with the volume of work focused on how intrasexual selection shapes sperm phenotype and number (Parker & Pizzari, 2010; Simmons & Fitzpatrick, 2012). Here, we extend the scope of research pertaining to non-sperm copulatory fluids and their effect on sperm competition by examining whether a non-sperm, non-*semen* reproductive fluid influences male reproductive success in competitive scenarios. By a "non-semen" copulatory fluid we refer to a fluid that is produced and stored separately from semen (sperm and non-sperm components of the seminal fluid), never packaged with sperm inside males, and not mixed with semen until inside the female reproductive tract. Male Japanese quail (*Coturnix japonica*) have a well-developed foam gland (also known as the "proctodeal gland" or "cloacal gland") that produces one such copulatory fluid (Klemm *et al.*, 1973). This large, red, external protuberance lies dorsal to the cloaca in sexually

mature males and secretes a viscous mucoprotein that males whip into a frothy, meringue-like foam (Figure 1.1 a-c) (Klemm *et al.*, 1973; Seiwert & Adkins-Regan, 1998). The exaggerated foam and foam-gland complex is novel to *Coturnix* males and there is debate about whether a homologous structure exists in other species (Klemm *et al.*, 1973; King, 1981). Like most birds, males of this genus have no other accessory gland and any seminal fluid is produced from the seminiferous tubules or epithelia of the testes or ductus deferens (Lake, 1981). During copulation, males introduce semen immediately followed by a large quantity of foam to females (Coil & Wetherbee, 1959). The foam and semen are never found together inside of males, and foam, unlike semen, is not required for fertilization (Ikeda & Taji, 1954; Marks & Lepore, 1965a; King, 1981). By examining whether this novel reproductive fluid influences sperm competition outcomes, we examine whether any reproductive fluid a male transfers to a female, not just sperm and seminal fluid comprising the semen, has the potential to evolve under sperm competition.

Sperm competition is most likely strong in Japanese quail. Systematic observations of this species' mating system are difficult in the wild as Japanese quail are cryptic and hard to observe. However, a study from a feral population in a semi-natural setting suggests that Japanese quail may form at least short-term pair relationships but also engage in mate switching and forced extra-pair copulations (Nichols, 1991). Research describing the mating system of wild populations of their closely related sister species, the European quail (*Coturnix coturnix*), indicates serial monogamy with opportunistic extra-pair copulations (Teijeiro *et al.*, 2003). Strong, indirect evidence also suggests that Japanese quail females mate multiply. Females are receptive to multiple males in lab settings and males can force copulations if females are not receptive (Adkins-Regan, 1995). Female Japanese quail can

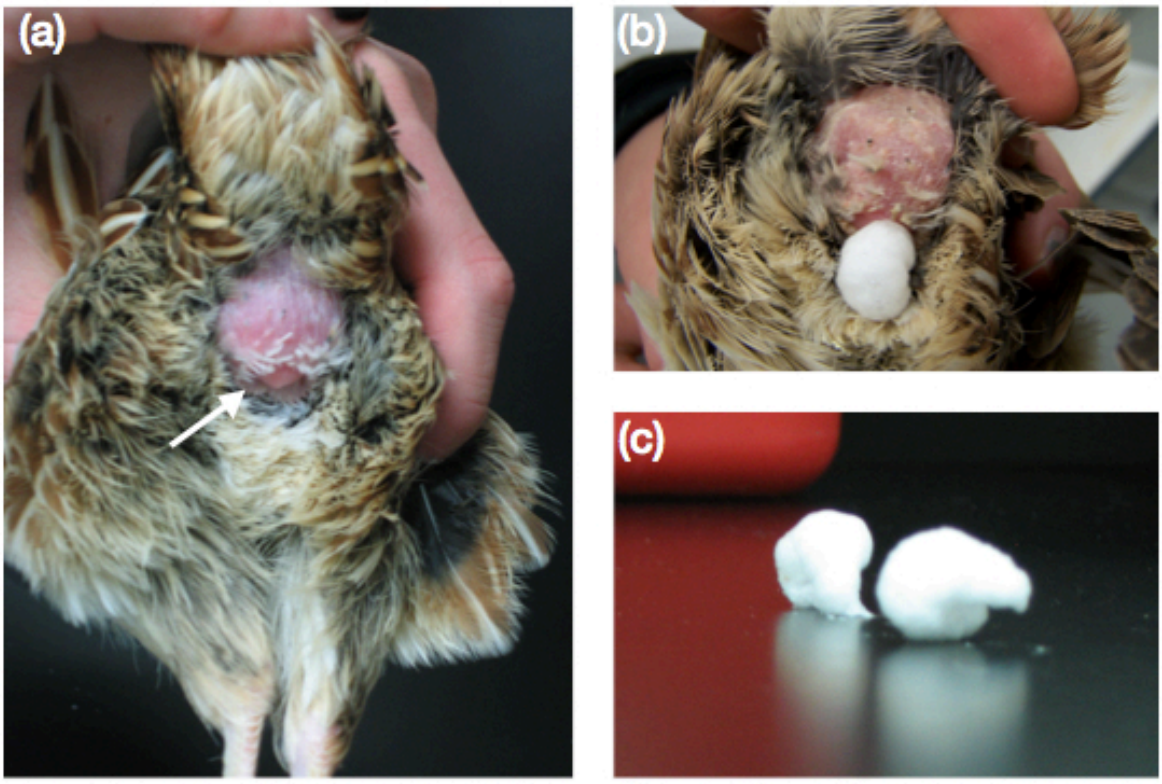

Figure 1.1 Foam gland system of male Japanese quail. a) The foam gland of the male Japanese quail. Arrow indicates cloacal opening. Feathers surrounding the foam gland are plucked. **b)** Manual expression of foam from foam gland. **c)** Foam manually expressed from two male Japanese quail. Photos by FR Finseth.

store sperm for up to 11 days, which would allow ejaculates from different males to overlap, even when matings occur on different days (Sittmann & Abplanalp, 1965; Birkhead & Fletcher, 1994b). Additionally, male Japanese quail have large testes for their mass and a high daily output of sperm, which are patterns seen in species with high levels of sperm competition (Clulow & Jones, 1982; Møller, 1991).

There are several ways foam could enhance a male's reproductive success. The presence of foam is known to improve a male's fertilizing efficiency at certain stages in a female's ovulatory cycle (Cheng, *et al.*, 1989a; Adkins-Regan, 1999), extend the duration of the period that a male can fertilize a female's set of eggs following a single insemination (Cheng, *et al.*, 1989a; Singh *et al.*, 2012) and increase sperm motility and viability *in vitro* (Cheng, *et al.*, 1989b; Cheng, *et al.*, 1989a; Singh, *et al.*, 2011b). Addition of foam to neat semen also disaggregates clumps of sperm (Singh, *et al.*, 2011b). The presence of foam may also help transport sperm into the oviduct during a natural mating, possibly by improving sperm motility (Singh *et al.*, 2012). For these reasons, we predict that a male's foam would have a positive effect on his own competitive ability, resulting in a concomitant negative effect on his rival's reproductive success. Earlier studies found mixed support when testing a similar hypothesis, but those studies either employed a foam removal manipulation confounded by an invasive surgery likely to interfere with copulation (Cheng, *et al.*, 1989b), mated competitor males from two different color strains to assign paternity (Cheng, *et al.*, 1989b), or did not allow sperm from two males to compete (Adkins-Regan, 1999). The experimental design of the current study improves upon previous work, as we utilize a non-invasive foam removal manipulation and determine paternity with polymorphic molecular markers.

To test the hypothesis that foam influences the reproductive success of males under sperm competition, we employ a within-subject design where females are mated to pairs of males, one of whose foam state is manipulated (Figure 1.2). Our main objective is to determine whether foam from one male reduces the ability of a competitor to compete for fertilizations of a female's set of eggs. Therefore, we predict that a male will have a fertilization advantage when his rival's foam is absent. We present results in support of this prediction. Additionally, we wish to understand how mating order affects competitive success with and without foam. We do so in the context of considering traits to be "defensive" or "offensive" (Parker, 1970; Parker 1984; Clark *et al.*, 1995), that is, by varying the order males are mated to females to disentangle whether foam confers a first or second male advantage (Figure 1.2). We find that first males have a competitive advantage independent of foam state.

Methods

Subjects

Japanese quail were hatched in late December 2009 from eggs purchased from Cumberland Game Farm. Birds were housed individually after four weeks (sexual maturity begins at 5-6 weeks). At the beginning of the experiment, all birds were three months old and housed on a 16:8 light:dark cycle, with lights on at 8:00. Prior to the experiment, males were screened for mating competency. Only those males that successfully copulated with a female at least once were used for the study. Mating trials were performed between 11:00 and 15:00, a time when females were likely to be receptive (Delville *et al.*, 1986). Mating

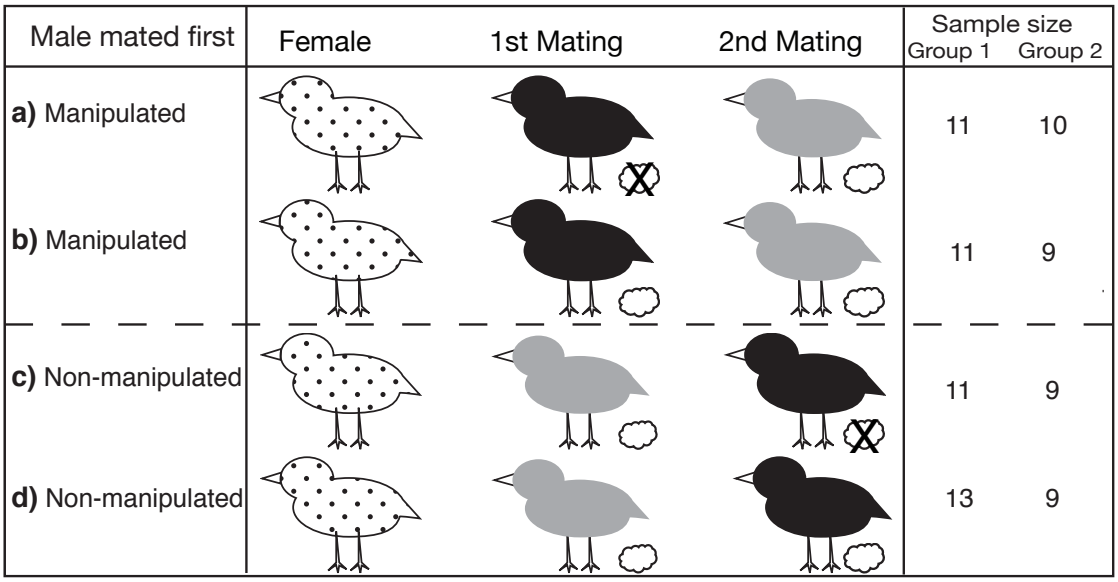

Figure 1.2 Experimental design. Birds were assigned to trios of one female (dots) and two males (black and gray). The males were designated as either a manipulated male (black), whose foam was sometimes removed, or a non-manipulated male (gray), whose foam was never removed. **a & b)** A manipulated male was mated on two separate occasions to a female first, once with his foam complement (white blob) removed (a) and once with it present (b). **c & d)** A manipulated male was mated second to a female, on two separate occasions. In one mating, the manipulated male's foam complement was removed (c) and once it was present (d). A non-manipulated male was predicted to father proportionately more of a clutch and fertilize more eggs when the manipulated male's foam was absent (a & c) when compared to when the manipulated male's foam was present (b & d). In all four treatments, the same female was mated to the same male pair. Individual trials for a given trio were separated by two weeks. The final sample sizes (i.e., the number of successful trials) for each treatment per group are given in the right column.

trials took place in a large testing cage (43.5 x 61 x 35 cm) and individuals were weighed prior to the start of the experiment.

Design

To examine whether foam influenced the outcome of sperm competition, we performed four different mating treatments on trios of birds where we manipulated both foam state (absence/presence of foam in one of the males) and mating order (first or second). A trio consisted of a female and a pair of males. Each male pair included a "manipulated" male, whose foam was removed in some treatments, and a "non-manipulated" male, whose foam was never removed. Trials where the manipulated male was mated first tested whether foam conferred a competitive advantage to a first male, and those where the manipulated male was mated second examined if foam was advantageous to second males. Treatments were as follows (Figure 1.2): a) manipulated male mated first, foam removed, b) manipulated male mated first, foam present, c) manipulated male mated second, foam removed, d) manipulated male mated second, foam present. As time of day can affect fertilization outcomes in this species due to differences in female receptivity and ovulation time, the same trio was mated at approximately the same time of day (+/- 30 minutes) across all four treatments (Birkhead, 1998). Each trio was subjected to one of the four treatments (Figure 1.2 a-d) once every two weeks, as female Japanese quail can store sperm for up to 11 days (Sittmann & Abplanalp, 1965; Birkhead & Fletcher, 1994b; Adkins-Regan, 1995). Treatments were randomized throughout the four-hour testing period and the order that the four treatments (Figure 1.2 a-d) were performed on a trio was counterbalanced.

The member of a male pair assigned to the manipulated role and the member assigned to the non-manipulated role were maintained throughout all four treatments in a single group (Figure 1.2 a-d). This design allowed for within-subject comparisons of a non-manipulated male across all four treatments. We attempted a total of 132 experimental trials on two groups of 16 and 17 trios, encompassing 35 females and 34 males. Only 83 of these trials were successful and included in our final statistical analysis. Prior to the start of the trials, males were paired according to mass, foam gland area (which approximates foam production and fertility (Mohan *et al.*, 2002; Biswas *et al.*, 2007)), and unique microsatellite profiles. Assigning appropriate male pairs was difficult given a limited number of birds and our pairing criteria. Therefore, for the second group, we kept the same male pairs but mated them to a new female and switched which member of the pair was in the manipulated or non-manipulated role. Thus, a male assigned the manipulated role in group one, became the non-manipulated male in group two. Males were therefore paired with the same competitor in both groups, but mated to a different female in each group. Trials for the two groups were run concurrently, but offset by one week, allowing males a week of recovery time between matings. We emphasize that what we term "groups" are not two independent groups of birds. Rather, the "groups" represent two sets of trials on one group of males (with their role assignments reversed in the two trial sets). To avoid pseudoreplication, all analyses account for the fact that repeated measures were taken on the same male pair (i.e., male pair is a random factor in our models).

Experimental methods

We manually expressed foam for our foam removal manipulation, which involved gentle squeezing of the foam gland and wiping away foam with a paper towel 45 seconds prior to mating (Figure 1.1 b). This manipulation does not interfere with a male's ability to copulate, ejaculate sperm, or inseminate a female under conditions similar to those in our study (Adkins-Regan, 1999). Neither sperm nor semen are found in manually expressed foam (Ikeda & Taji, 1954; King, 1981). This manipulation significantly reduces foam for 20 minutes and it takes 40 minutes for a male to regenerate a full foam complement (Adkins-Regan, 1999). Additionally, when females are mated to males with manually expressed foam, no foam is evident in the female's cloaca (Adkins-Regan, 1999). Previous work showed that in non-competitive matings at the time of day when our mating trials were held, manual expression of foam does not affect 1) the likelihood of obtaining any fertilizations, 2) the number of eggs fertilized by a male, or 3) a female's receptivity to a second male (Adkins-Regan, 1999). One drawback of this technique is that manual expression may leave trace amounts of foam in the foam gland. Therefore, this manipulation is technically a significant reduction, not a total elimination, of foam. However, for clarity, we refer to males with manually expressed foam complements as "foam removed" or "without foam" throughout the paper. If a male's foam was not removed, he was handled in a similar manner (briefly held on his back and touched around the cloacal area 45 seconds prior to mating) to control for handling during foam removal.

For all trials, a female was mated with a pair of males in succession. Females were placed in the testing cage and allowed to acclimate for two minutes. We then introduced the first male into the testing cage and allowed him four minutes to copulate with the female. The first male was removed immediately following a successful copulation. The

female was rested for four minutes before introducing a second male to the testing cage. Again the second male was given four minutes to mate with the female, and removed immediately following a successful copulation. A successful copulation was classified by close observation of a characteristic suite of male behaviors (cloacal contact followed by immediate dismount, cessation of copulatory attempt, fluffing and/or strutting and/or crowing) to ensure that females received a single, successful insemination. Previous work showed that observation of this characteristic suite of behaviors is followed by a successful insemination in a minimum of 90.6% of observations (Adkins-Regan, 1995). Additionally, we confirmed a successful insemination by observing a female's cloacal region post-mating (*e.g.*, looking for foam presence when appropriate). We chose to use natural matings rather than artificial insemination because prior work in Japanese quail shows that 1) artificial insemination of less than three inseminations, as with our design, results in very low proportions of fertilized eggs (4 - 36 %) (Wilson *et al.*, 1961; Ogasawara & Huang, 1963; Marks & Lepore, 1965b), 2) to achieve fertility levels seen in natural matings, semen must be placed in the vagina or uterus, which is much farther up the reproductive tract than where males deposit foam (Wentworth & Mellen, 1963; McFarquhar & Lake, 1964; Marks & Lepore, 1965a; Cheng, *et al.*, 1989a) and 3) artificially placed foam has less of an impact on fertility than naturally placed foam (Cheng, *et al.*, 1989a).

In this species, females lay eggs in the afternoon and can store sperm for a maximum of 11 days (Sittmann & Abplanalp, 1965; Birkhead & Fletcher, 1994b; Adkins-Regan, 1995). We collected eggs in the morning for the 11 days after mating. The egg collected the first morning following mating was discarded, as this egg is never fertilized because it is past the point of fertilization in the female's reproductive tract at the time of

mating (Adkins-Regan, 1995). Eggs were stored at 7.2 °C for up to six days prior to incubation at 37.5°C and approximately 30% relative humidity for one week. Eggs were then broken and examined for embryonic development. Eggs were classified as unfertilized (clear albumen, unmixed yolk), normal embryos (properly developed 7-day embryo), or failed embryos (evidence of stunted development either from milky yolk/albumen, appearance of blood vessels only, or abnormal embryonic appearance). Tissue samples were collected from eggs with both normal and failed embryos and frozen at -80 °C for later use. Eggs with both normal and failed embryos were considered "fertilized". The set of fertilized eggs produced by a single female from one mating trial where she successfully copulated with both males is referred to as a "clutch" throughout.

Paternity analysis

Blood was collected from adult birds prior to the start of the experiment by alar (wing) venipuncture and frozen at -80 °C until DNA extraction. We extracted DNA from the blood samples of 88 adult birds using the DNEasy® Blood and Tissue Kit (QIAGEN®) following the manufacturer's protocol for purification of total DNA from nucleated blood cells. We extracted DNA from 394 embryos using the Agencourt® DNAdvance™ Genomic DNA Isolation Kit (Beckman Coulter®) following the manufacturer's instructions, with the exception that we did half reactions and used approximately 5 mg of tissue.

All individuals (394 embryos, 88 adults) were genotyped for six polymorphic loci originally isolated from Japanese quail (Table S1.1; Kayang *et al.*, 2000; 2002). In order to pair males with sufficiently distinct microsatellite profiles to aid in paternity assignment, adults were genotyped prior to the start of the experiment. To facilitate multiplexing and

improve peak calling, we added a long M13 sequence tag (5' CGA GTT TTC CCA GTC ACG AC 3') to the 5' end of forward primers as described previously (Boutin-Ganache *et al.*, 2001). Additionally, we pooled the six primers into two different mixes for polymerase chain reaction (PCR), with each mix having a different colored fluorescently labeled M13 primer (Table S1.1). Each 10 μL PCR contained 1 μL of template DNA, 2 mM $MgCl_2$, 10 μM dNTPs, 1 μL 10X Platinum® *Taq* PCR Buffer (Invitrogen®), 1 U Platinum® *Taq* DNA polymerase (Invitrogen®), 0.00025 μM of each of three M13-tailed forward primers, 0.002 μM of each of three reverse primers, and 0.0675 μM of a fluorescently labeled M13-primers (Applied Biosystems®). Fluorescent tags and the three primer pairs that went into each of the two mixes are given in Table S1.1. The thermal profile for PCRs was as follows: preheat at 95 °C for 4 min, denature at 94 °C for 50 s, anneal at 55 °C for 50 s, extend at 72 °C for 90 s. PCRs were performed for 30 cycles with a final extension at 72 °C for 10 min (Thermo Scientific®, Hybaid Px2 Thermal Cycler). The PCR products were diluted and run on an ABI PRISM® 3100 Genetic Analyzer (Applied Biosystems®). We used the software GeneMapper® to view and score alleles (Applied Biosystems®).

We assigned paternity using the program Cervus in combination with manual exclusion of one father (Marshall *et al.*, 1998). Cervus is a maximum likelihood program that determines confidence levels for paternity assignments via simulations. We performed 10,000 simulations of parentage analysis using the genotypes of all potential parents. Based on our simulations, the probability of our data to exclude a second parent was 0.93. We assigned a known mother and two candidate fathers to each embryo and asked Cervus to return the most likely father at the 95% confidence level. After initially assigning paternity using likelihood in Cervus, we confirmed all assignments by examining

whether one of the candidate males was excluded as a potential father by observation of at least one mismatched allele at one or more loci. We also used Cervus to calculate observed heterozygosity (H_o), expected heterozygosity (H_e), polymorphism information content (PIC), and test for the frequency of null alleles in the parental population (Table S1.1). After initially genotyping 88 adults, we chose 35 females and 34 males to make experimental trios based on their masses, foam gland areas (in males), mating competency, and the ability to discriminate their microsatellite profiles.

We genotyped 394 embryos and 88 adults from all successful trials for six microsatellite loci. One locus (GUJ0041; Table S1.1) revealed significant frequencies of null alleles and was removed from all further analyses. The remaining five loci were in Hardy-Weinberg equilibrium in the adults and displayed lower than 10% frequency of null alleles (Table S1.1). Two embryos did not match their known mother at one locus and an additional two embryos could not be confidently assigned one of the two candidate fathers. These four embryos were removed from the study. For the remaining 390 of the 394 embryos (99% of all fertilized eggs), Cervus assigned paternity at the 95% confidence level. We were able to manually confirm these 390 assignments, as the excluded male mismatched at least one allele of each embryo, given its known mother, while the assigned parents and embryos had no mismatched alleles in any case.

Data analysis

We attempted 132 trials. Trials where the female did not successfully mate with both males (42 trials), when it was apparent that foam was not transferred properly (*e.g.*, we observed foam or semen on the cloacal lip of the female rather than inside her cloaca; 3

trials), or when a male was mistakenly repeated in a pair (4 trials) were removed from the study. Unless otherwise stated, our final sample size was 83 distributed among the treatments and groups as in Figure 1.2.

To evaluate how foam affected sperm competition success, we analyzed whether the following differed according to foam state: 1) the proportion of eggs fertilized by a non-manipulated male, 2) the probability of either male fertilizing any eggs, and 3) the duration of fertility. For the first two objectives, generalized linear mixed models (GLMMs) fitted to a binomial distribution with a logit link function were used to determine the fixed effects of foam (2 levels: presence/absence), mating order (2 levels: first/second), and the interaction of foam and mating order on the proportion of eggs fertilized by a non-manipulated male. For objective 1, the number of eggs fertilized by a non-manipulated male for one trial was the number of "successes", while all other fertile eggs from a clutch (*i.e.,* the number fertilized by the manipulated male) were "failures". We ran this analysis on the full dataset (N=83) and the subset of trials with at least one fertile egg (N=69). For the second objective, we classed each mating trial as either "yes" or "no" according to whether a non-manipulated male fertilized any of a female's set of eggs for that trial. We repeated this second analysis to examine whether the proportion of eggs fertilized by a manipulated male differed according to foam status. For duration of fertility, linear mixed models analyzed the effect of foam, mating order, and the interaction of foam and mating order on the last day a manipulated male fertilized an egg. Only trials where a manipulated male fertilized an egg (N=39) were retained for the duration of fertility analysis.

For all analyses, model selection was performed by comparison of AIC values generated from models fit by maximum likelihood. In all models, male ID nested within

male pair was included as a random factor to account for the within-subject design and the fact that male pairs were repeated in the two the groups. To determine significance of fixed effects and generate *P*-values, targeted likelihood ratio tests (LRTs) comparing models with and without the term of interest were conducted. Parameters of best models were estimated with maximum likelihood of binomially distributed response variables and restricted maximum likelihood for normally distributed response variables (*i.e.*, duration of fertility analyses). Parameter estimates and standard errors are given in Table S1.2. All analyses were implemented in the lme4 package (Bates *et al.*, 2012) of R Version 2.15.2 (R Core Team 2012). Figures were made in the ggplot2 package of R (Wickham, 2009). We generated 95% confidence intervals by performing 10,000 bootstrap resamplings of the mean without assuming normality implemented in the Hmisc package of R (Harrell, 2012). To display within-group patterns, our data are presented graphically separated into groups (Figures 1.3, 1.4). Table S1.3 presents overall and group-specific proportions or means and standard errors.

Results

In 83 out of 132 trials (62.9%), both males successfully copulated with the female. Successful trials are distributed among treatments and groups as in Figure 1.2. In 69 of 83 successful trials, females laid at least one fertile egg (83.1% of successful trials) and 22 matings resulted in mixed paternity clutches (26.5% of successful trials). Fourteen of 42 foam-present (33.3%) and 8 of 41 foam-removed (19.5%) trials resulted in mixed paternity.

Foam and sperm competition outcomes

The best model explaining the proportion of offspring fathered by the non-manipulated male included the fixed effects of foam presence and mating order (Table 1.1a). For all models, parameter estimates for terms in best models are given in Table S1.2. The proportion of eggs fertilized by a non-manipulated male was significantly higher when the manipulated male's foam was absent (LRT of mating order versus foam + mating order model; $\chi^2_{[1]} = 16.375$, $P = 5.19 * 10^{-5}$; Figure 1.3a; Table S1.2a). The non-manipulated males also fertilized proportionately more of a clutch when mated first (LRT of foam versus foam + mating order model: $\chi^2_{[1]} = 28.931$, $P = 7.5 * 10^{-8}$; Figure 1.3a; Table S 1.2a). Figure 1.3 also presents the number of fertilized eggs obtained by the non-manipulated (1.3b) and manipulated males (1.3c) for visualization of how the data were distributed in terms of absolute numbers of fertilizations achieved. Our results were the same for models run on the full dataset (N=83) and the subset of trials with at least one fertile egg (N=69).

Foam was included in the best models explaining the probability of either male gaining any fertilizations, but only had a significant effect when considering the manipulated male (Table 1.1b and 1.1c). For non-manipulated males, the best model included foam presence and mating order (Table 1.1b). Mating first increased the probability of a non-manipulated male gaining any fertilizations, (LRT of foam versus foam + mating order model: $\chi^2_{[1]} = 4.2175$, $P = 0.040$; Figure 1.4a and 1.4b; Table S1.2b), but foam presence did not (LRT of mating order versus foam + mating order model: $\chi^2_{[1]} = 2.461$, $P = 0.117$; Figure 1.4a and 1.4b; Table S1.2b). Foam presence was the only term in the best model explaining the probability that a manipulated male fertilized any eggs and had a significant positive effect (LRT of mating order versus foam + mating order model:

Table 1.1: Model selection process for (a) proportion of a clutch fertilized by a non-manipulated male, probability of a (b) non-manipulated and (c) manipulated male fertilizing any eggs, and (d) duration of fertility. "Foam" refers to foam state. Best models are shown in bold.

Terms	k^*	AIC	Δ AIC†
(a) Proportion of clutch fertilized by a non-manipulated male:			
Foam + Mating order + Foam * Mating order	6	238.7	1.7
Foam + Mating order	**5**	**249.0**	**0.0**
Foam	4	275.9	26.9
Mating order	4	263.3	14.3
Null	3	290.7	41.7
(b) Probability of a non-manipulated male fertilizing any eggs:			
Foam + Mating order + Foam * Mating order	6	110.7	0.4
Foam + Mating order	**5**	**110.3**	**0.0**
Foam	4	112.9	2.6
Mating order	4	110.8	0.5
Null	3	112.9	2.6
(c) Probability of a manipulated male fertilizing any eggs:			
Foam + Mating order + Foam * Mating order	6	116.7	2.6
Foam + Mating order	5	115.9	1.8
Foam	**4**	**114.1**	**0.0**
Mating order	4	122.3	8.2
Null	3	120.5	6.4
(d) Duration of fertility:			
Foam + Mating order + Foam * Mating order	7	168.7	1.3
Foam + Mating order	6	169.3	1.9
Foam	5	168.1	0.7
Mating order	5	168.7	1.3
Null	**4**	**167.4**	**0.0**

$^*k$ represents number of parameters.

†Δ AIC is calculated from subtracting a particular model's AIC value from the best-fitting model for that comparison.

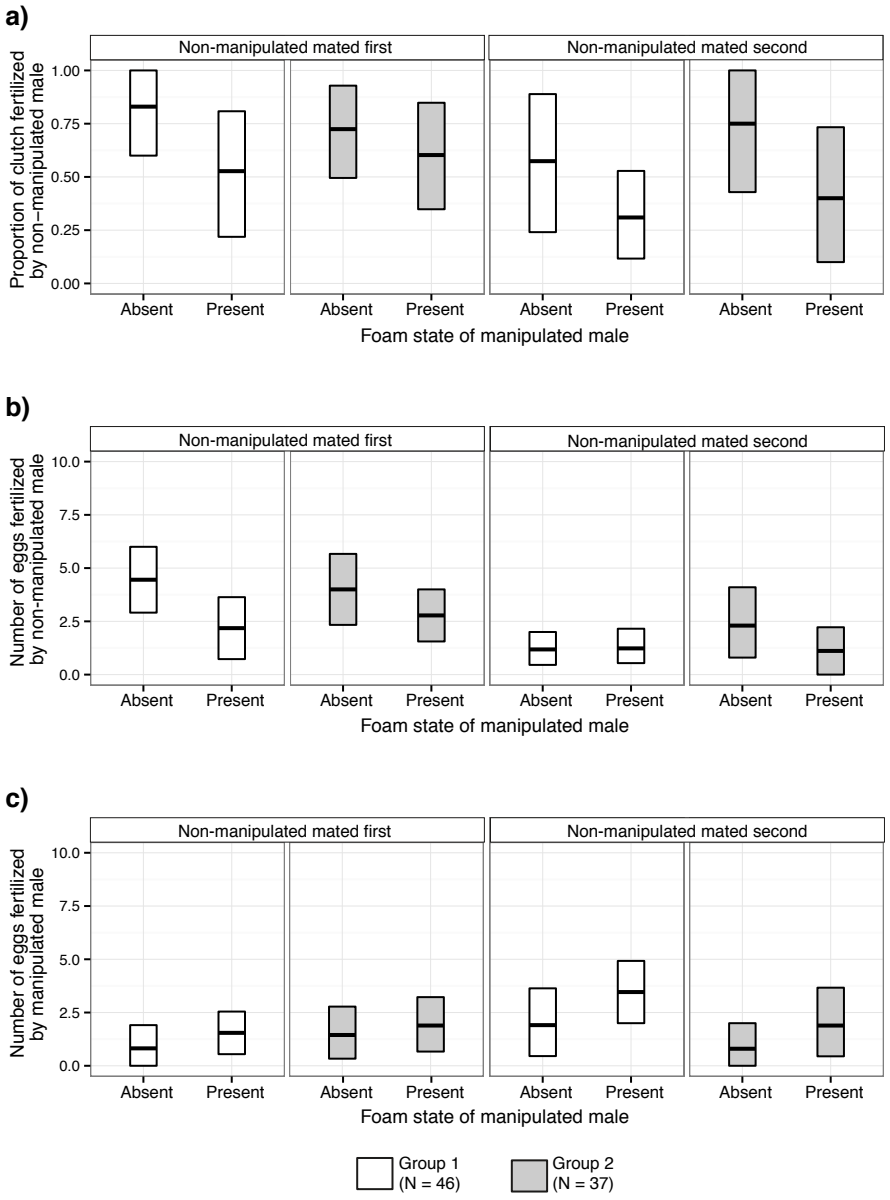

Figure 1.3. A non-manipulated male fathers a larger proportion of a clutch when his competitor's foam is absent. a) The proportion of a clutch fertilized by a non-manipulated male and the number of eggs fertilized by **b)** a non-manipulated male or **c)** a manipulated male when the manipulated male's foam was absent or present. Foam absence ($\chi^2_{[1]} = 16.375$, $P = 5.19 * 10^{-5}$) and mating first ($\chi^2_{[1]} = 28.931$, $P = 7.5 * 10^{-8}$) significantly increased the proportion of a clutch fertilized by a non-manipulated male. Thick black lines represent means, boxes represent 95% confidence intervals. Results from groups 1 and 2 are displayed in white and gray boxes, respectively. Trials where the non-manipulated male was mated first are presented in the two left panels, and those where the non-manipulated was mated second are in the two right panels. "N" indicates number of successful trials in each group. Significance was determined via targeted LRTs of GLMMs fitted to binomial distributions.

$\chi^2_{[1]} = 8.450$, $P = 0.00365$; Table 1.1c; Figure 1.4c and 1.4d; Table S1.2c). Mating order did not significantly affect the probability a manipulated male was able to fertilize any eggs (LRT of foam versus foam + mating order model: $\chi^2_{[1]} = 0.202$, $P = 0.653$).

We found no evidence that the presence of a manipulated male's foam, mating order, or the interaction of mating order and foam extended his duration of fertility. The null model fit the data best and neither foam (LRT of mating order versus foam + mating order model: $\chi^2_{[1]} = 0.628$, $P = 0.428$) nor mating order (LRT of foam versus foam + mating order model: $\chi^2_{[1]} = 0.037$, $P = 0.847$) were significant (Table 1.1d; Table S1.2d). Interaction terms were not retained in any of the best models (Table 1.1 a-d) and were not significant ($P > 0.05$ in all cases).

Discussion

Foam and sperm are required to influence sperm competition outcomes

In the present study, foam, a non-semen copulatory fluid, improved a male's competitive reproductive success at the expense of his rival's. Two main lines of evidence allow us to draw this conclusion: 1) the proportion of a clutch fertilized by a non-manipulated male increased when his competitor's foam was absent (Figure 1.3a; Table 1.1a; Table S1.2a), and 2) the chance that a manipulated male fertilized any eggs was greater when his foam was present (Figure 1.4c and 4d; Table 1.1c; Table S1.2c). Our results provide no evidence that foam differentially conferred first or second males an advantage, as interaction terms were not included in any best models (Table 1.1a-d). However, mating order did influence fertilization outcomes independent of foam. When a non-manipulated male mated in the first position, he fertilized a larger share of eggs

(Figure 1.3a; Table 1.1a) and had a higher probability of obtaining any fertilizations (Figure 1.4ab, Table 1.1b; Table S1.2a and S2b).

In non-competitive matings, Adkins-Regan (1999) found a similar increase in the probability a male gained any fertilizations when his foam complement was present. In the previous study, the effect was driven by matings that took place when females had hard-shelled eggs present in their uteri (18:00 to 19:00), though matings were performed throughout the day. This finding was consistent with a hypothesis originally suggested by Cheng *et al.* (1989 a,b), whereby foam minimizes sperm loss by suspending sperm in the proctodeum of females away from the path of the egg until after oviposition. Here, we performed competitive matings to test a sperm competition hypothesis. Our matings were conducted from 11:00 to 15:00, outside of the time of the ovulatory cycle immediately preceding oviposition. Thus, our results demonstrate fertility benefits of foam under competition beyond prevention of sperm loss from oviposition of a hard-shelled egg.

Two earlier studies explicitly examined the role of foam in sperm competition. Cheng *et al.* (1989b) found that males with foam fathered 99% of progeny when competed against males without foam. While we demonstrate a positive influence of foam on sperm competition, our results are not so dramatic (Figure 1.3a, Table S1.2a). The foam removal treatment used by Cheng et al. (1989b) involved cauterization of the foam gland, which often results in damage to the cloacal sphincter muscle. This muscle is essential for cloacal contact, ejaculation, and sperm transfer (King, 1981; Seiwert & Adkins-Regan, 1998). Our protocol involved a foam manipulation (manual expression) that is likely a better mimic of how foam affects competitive reproductive success in natural copulations. A second earlier study found no support for the hypothesis that foam plays a role in sperm competition after

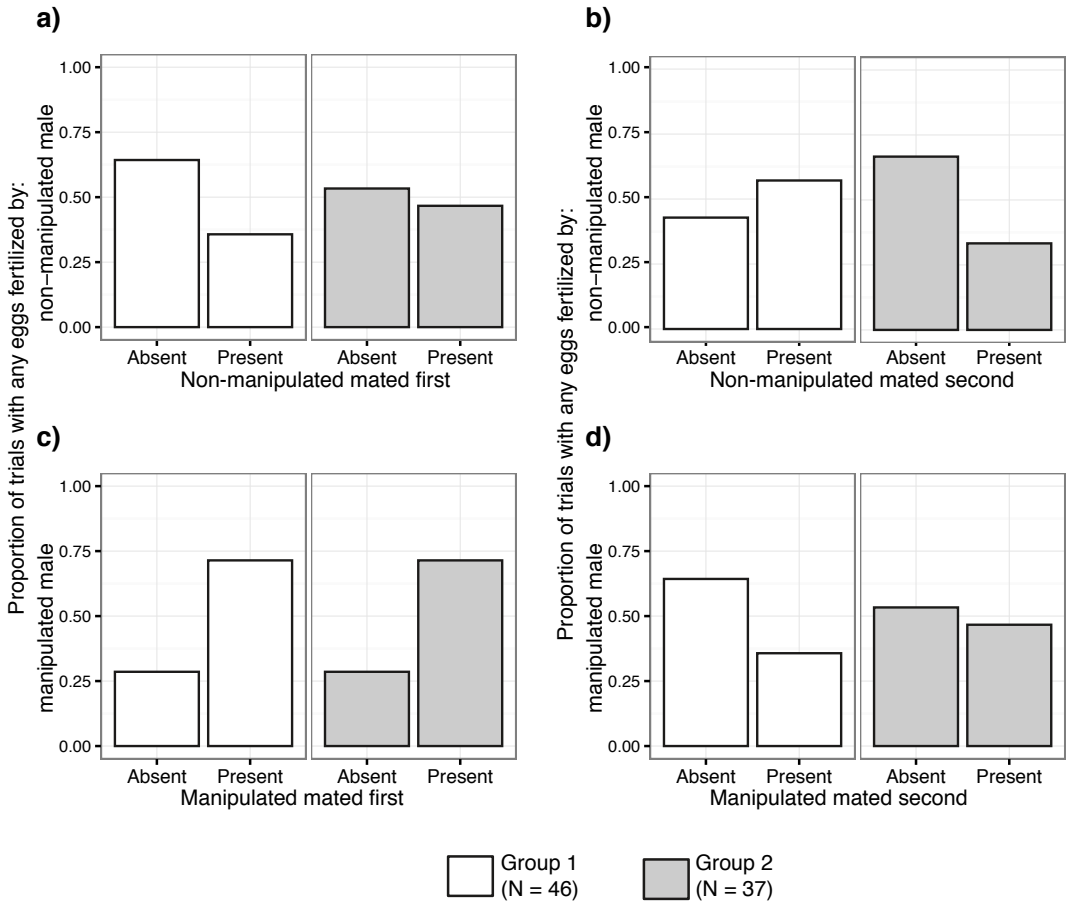

Figure 1.4 A manipulated male is more likely to fertilize any eggs when his foam is present. Proportion of trials where the **a & b)** non-manipulated male or **c & d)** manipulated male fertilized any eggs when the non-manipulated (a & c) or manipulated (b & d) male was mated first. Foam presence increased the likelihood a manipulated male obtained any fertilizations ($\chi^2_{[1]} = 8.450$, $P = 0.00365$; c & d), while non-manipulated males had a higher probability of obtaining any fertilizations when mated first ($\chi^2_{[1]} = 4.2175$, $P = 0.040$; a & b). Results from group 1 are white and those from group 2 are gray. Significance was determined via targeted LRTs of GLMMs fitted to binomial distributions.

experimentally introducing foam alone into a female's cloaca and determining its impact on a mating male's fertilization success (Adkins-Regan, 1999). The experimental design in this study was similar to the present one, with the important exception that foam was introduced to females by either 1) manually depositing foam in the cloaca with a blunt-tipped syringe or 2) allowing females to mate with "foam donor" males who transferred seminal fluids and foam to females, but no sperm due to photic castration with hormone replacement (Adkins-Regan, 1999). Here, we varied the foam state of one male, yet allowed both male's sperm to compete against each other, and found a significant role of foam in sperm competition. Foam alone, therefore, does not directly negatively affect a competitor's sperm. Rather, the critical factor affecting sperm competition is that a competing male transfers both his foam and sperm to a female. Thus, the influence of foam on sperm competition outcomes likely depends on the positive effects of foam from a given male on his own sperm. Foam from a focal male may increase his fertilizing efficiency or the likelihood sperm gets stored, with associated negative effects on a rival.

While foam is not required for fertilization and never mixed with sperm in males, we conclude that any effect of foam under sperm competition is mediated through the interaction of a male's own foam and sperm. Studies of seminal fluid proteins in insects have come to similar conclusions. In *D. melanogaster*, seminal fluid proteins are mediators of sperm displacement, as seminal fluid alone from a second male reduces a previously mated female's fertility (Harshman & Prout, 1994; Gilchrist & Partridge, 1995; Prout & Clark 2000; reviewed in Ram & Wolfner, 2007). However, the displacement effect is greater when second males transfer both sperm and seminal fluid proteins, suggesting that

the interaction of sperm and seminal fluid proteins is required to observe the full extent of sperm displacement (Gilchrist & Partridge, 1995; reviewed in Ram & Wolfner, 2007). *D. melanogaster* females also exhibit a short- and long-term post-mating response that includes an increase in egg production, ovulation, and egg deposition and a reduction in receptivity to re-mating (reviewed in Chapman & Davies, 2004; Ram & Wolfner, 2007; Sirot *et al.*, 2009)). A network of interacting seminal fluid proteins, including a well-studied protein called sex peptide, modulate the long-term post-mating response in females, but the response only occurs when sperm are present (Ram & Wolfner, 2005; 2009). Sex peptide, critical to the long-lasting nature of a female's post-mating response, binds to sperm, which stabilizes it and allows for slow release of the peptide over several days (Peng *et al.*, 2005). When females have sperm bound with sex-peptide in storage, they are likely to display increased egg-laying and reject mating attempts by males, thereby indirectly affecting sperm competition (Peng *et al.*, 2005; reviewed in Sirot *et al.*, 2009).

Although most species of birds, including other domesticated galliform species, show last male precedence (Birkhead, 1998), we found that first males had a fertilization advantage. Last male precedence arises because birds, including Japanese quail, experience passive sperm loss, meaning that sperm are continuously lost from the female's sperm storage tubules (Birkhead & Fletcher, 1994b; Birkhead & Møller, 1998; Birkhead, 1998). The second of two inseminations will father more offspring simply because fewer of the second male's sperm have been passively lost. The key determinant of the probability of fertilization is the interval between matings—the longer the time interval, the greater the differential paternity success of the second male (Birkhead & Møller, 1998; Birkhead, 1998). In our study, we wanted to maximize the chance that sperm and foam from two

males interacted in the female's reproductive tract and kept the interval between matings short (four minutes). The short time interval may explain the discrepancy in precedence patterns between our results and previous studies. An alternative explanation is that Japanese quail have different sperm precedence patterns than other avian species. Another recent study is consistent with this hypothesis, finding a lack of second male precedence (though no evidence that first males were favored) in Japanese quail (Singh, *et al.*, 2011a).

Potential mechanistic explanations for the function of foam in competitive scenarios

Synthesizing the present study's results with previous work, we conclude that Japanese quail males must transfer both foam and sperm to influence sperm competition outcomes (Adkins-Regan, 1999). Below, we discuss potential mechanistic explanations for how foam may interact with sperm to affect competitive reproductive success. These include: i) increased number of stored sperm, ii) enhanced sperm fertilizing efficiency, and iii) greater sperm retention. These mechanisms are not exclusive and may act synergistically.

i) Increased number of stored sperm

In birds, paternity under sperm competition is largely determined by the relative numbers of sperm from rival males in storage at the time of fertilization (Birkhead, 1998). Therefore, males with adaptations that increase the number of live, functional sperm that reach and remain in storage have a competitive advantage. Evidence from other systems indicates that the presence of a seminal fluid protein can affect the amount of sperm in storage, which can have implications for sperm competition. In *D. melanogaster,* the seminal fluid protein Acp 36DE is required for females to properly store sperm (Neubaum

& Wolfner, 1999; Chapman *et al.*, 2000). Acp 36DE null males end up losing in sperm competition because their sperm are poorly stored (Chapman *et al.*, 2000). It is possible that foam may be mediating a similar phenomenon in Japanese quail. Our results and a previous study show that foam improves the fertilization success of a focal (*i.e.*, manipulated) male by increasing the likelihood he fertilizes any eggs (Adkins-Regan, 1999). This is consistent with the idea that foam may determine whether a male gets any of his own sperm into the sperm storage tubules. Japanese quail females artificially inseminated with sperm and foam have a higher number of spermatozoa trapped between perivitelline layers (correlated with the number of sperm in storage) than females solely inseminated with sperm (Brillard & Antoine, 1990; Singh *et al.*, 2012). Taken together, foam may influence whether sperm are properly transferred to the female reproductive tract and/or make it to the sperm storage tubules in sufficient numbers to fertilize any ova. Moreover, improved storage of one male's sperm in the presence of his foam could be responsible for the observed decrease in the relative fertility of a rival.

ii) Enhanced sperm fertilizing efficiency

The fertilizing efficiency of an ejaculate determines the outcome of sperm competition after controlling for sperm number (Snook, 2005). Foam could improve competitive fertilization success through increasing the fertilizing efficiency of sperm. For example, foam could extend sperm longevity, allowing males with foam to fertilize eggs for a longer period of time (Snook, 2005; Pizzari & Parker, 2009). We found no evidence that foam extended the duration of a male's fertility, as the last day a manipulated male fertilized an egg did not differ according to foam state. This result contrasts with data from two previous studies (Cheng, *et al.*, 1989a; Singh *et al.*, 2012). In one study, males with

cauterized foam glands were only able to fertilize eggs for a maximum of 5 days, while males with sham-operated foam glands or cauterized foam glands plus manually deposited foam were able to fertilize eggs for 7-8 days (Cheng, *et al.*, 1989a). A second study found that adding foam to semen prior to depositing it in a female's vagina lengthens the fertile period from 2 to 5 days (Singh *et al.*, 2012). We observed that foam did not change the duration of fertility, but at least one male had foam present in all of our trials. Synthesizing these results suggests that foam may extend the amount of time sperm are able to fertilize eggs, as long as it is present in some form—either foam from the focal or another male. The idea that sperm quality improves in the presence of any foam, regardless of the source, may be a general phenomenon not restricted to sperm longevity. For example, foam enhanced sperm motility, regardless of whether foam was self-derived or from a different male (Cheng, *et al.*, 1989b).

Foam may enhance sperm fertilizing efficiency in other ways besides longevity. Foam has already been shown to increase motility (Cheng, *et al.*, 1989b; Biswas *et al.*, 2010; Singh, *et al.*, 2011b), disassemble clumped sperm (Singh, *et al.*, 2011b), improve sperm transport (Singh *et al.*, 2012), and possibly aerate sperm (Cheng, *et al.*, 1989b). Any one of these effects could improve a male's chances in sperm competition through increasing the likelihood sperm make it into the sperm storage tubules in sufficient numbers to outcompete rivals. Additionally, once sperm make it to the sperm storage tubules, more mobile sperm may have an advantage. In domestic fowl, low mobility ejaculates are more likely to fertilize the first egg after an insemination. Over subsequent days sperm are in storage, however, that advantage quickly disappears and high mobility ejaculates prevail, even when at a numerical disadvantage (Pizzari *et al.*, 2008). This may be explained by the

fact that more mobile ejaculates are lost from sperm storage tubules at a significantly slower rate (Froman *et al.*, 2002). Also in chickens, males with high sperm mobility outcompete males with low sperm mobility, all else being equal (Birkhead *et al.*, 1999). Recent work suggests foam may provide fuel to sperm in the form of lactate, possibly increasing longevity, metabolism and motility of sperm (Singh, *et al.*, 2011b).

iii) Greater sperm retention

The sperm retention hypothesis was developed by Hinton (1964) and originally applied to copulatory plugs, suggesting that copulatory plugs may prevent the loss of sperm from the female's reproductive tract (reviewed by Simmons, 2001). Foam is not a copulatory plug, but the increased likelihood of a manipulated male fertilizing any eggs when his foam was present could be due to foam minimizing sperm loss. If foam does improve sperm retention, it could be very advantageous as sperm can be lost from the female reproductive tract in several ways including removal through oviposition, leakage, or ejection by females (Cheng, *et al.*, 1989b; Adkins-Regan, 1999). That foam is not eliminated after oviposition (Cheng, *et al.*, 1989b), that female ejection has only a small effect on the probability of fertilization when males have foam (Adkins-Regan, 1995), and that Japanese quail have lower rates of instantaneous sperm loss than other species (Birkhead & Fletcher, 1994a) all suggest this may be a plausible mechanistic explanation.

Conclusions and remaining questions

We report that foam, a non-semen copulatory fluid, influences sperm competition outcomes by improving the likelihood that a male is able to fertilize any eggs, while simultaneously decreasing the proportion of a clutch fertilized by his rival. We compare

our results with the effect of foam in non-competitive matings from previous work and draw two important conclusions. First, adding sperm competition enhances the impact of a copulatory fluid. Under sperm competition, foam increases the probability that a male gains fertilizations beyond the narrow time window when foam affects fertility in non-competitive matings. Second, the impact of foam on sperm competition arises from positive effects of foam on a given male's sperm. The finding that sperm competition outcomes are mediated through the interaction of sperm and a novel, extra-semen copulatory fluid suggests that males can evolve complex strategies to gain a fertilization advantage at the expense of rivals.

While foam currently plays a role in sperm competition, we do not know if sperm competition is the reason the novel foam gland system originally evolved or if it is currently subject to post-copulatory sexual selection. Clues from domesticated birds suggest that there may be variation in foam production for sexual selection to act upon. Foam gland area and foam weight vary among domesticated Japanese quail with the most fertile males having the largest foam glands (Biswas *et al.*, 2007). As the species' history likely included frequent mate switching and forced copulations (Nichols, 1991; Teijeiro *et al.*, 2003), males with larger foam glands and/or males who make larger volumes of foam may have had a selective advantage. While this hypothesis has not been explicitly tested, there is some evidence that foam gland size may be important in a pre-copulatory context. A recent study suggests males with larger foam glands are more likely to secure successful copulations, although it is not clear if this effect is due to male-male competition, female choice, or simple differences in body size or hormonal state of males with larger glands (Singh *et al.*, 2012).

The presence of foam clearly improves a male Japanese quail's ability to secure fertilizations in both competitive and non-competitive scenarios (present study; Cheng, *et al.*, 1989b; Cheng, *et al.*, 1989a; Adkins-Regan, 1999; Singh, *et al.*, 2011b; Singh *et al.*, 2012). What remains unclear is why the foam evolved apart from the seminal fluid and why the foam-gland system is unique to *Coturnix* males. In addition to the foam gland, Japanese quail possess several other exceptional reproductive traits. Among domesticated galliform species, Japanese quail display the shortest duration of fertility following a single intravaginal insemination (Sittmann & Abplanalp, 1965). The sperm of Japanese quail males possess extremely large midpieces (64-74% of the overall length of the sperm) with an exceedingly high number of mitochondria (>1,400) per sperm (Korn *et al.*, 2000). Further, female Japanese quail are unusual in that they oviposit in the late afternoon or early evening (Wilson & Huang, 1962). Studies investigating how foam interacts with these exceptional characteristics of Japanese quail reproductive biology will offer insight into the diverse and complex strategies males can evolve under sperm competition.

Acknowledgements

We would like to thank S. Bogdanowicz for laboratory assistance, B. Miner for statistical advice, A. Ransler, N. Baran, and C. Schweitzer for help with experimental manipulations, and T. van Deusen, P. Smith, L. Vann and S. Martin for quail care. Thanks to the Harrison lab and four anonymous reviewers for comments on earlier versions of the manuscript. Funding for this project was provided by NSF DDIG DEB-1010757 to FRF and RGH, a P.E.O. Scholar award to FRF, and a Cornell University administrative supplement to EAR. The authors have no conflicts of interest.

References

Adkins-Regan, E. 1999. Foam produced by male *Coturnix* quail: What is its function? *Auk* **116**: 184–193.

Adkins-Regan, E. 1995. Predictors of fertilization in the Japanese quail, *Coturnix japonica*. *Animal Behaviour* **50**: 1405–1415.

Alonzo, S.H. & Pizzari, T. 2010. Male fecundity stimulation: conflict and cooperation within and between the sexes: model analyses and coevolutionary dynamics. *Am Nat* **175**: 174–185.

Andersson, M.B. 1994. *Sexual selection*. Princeton University Press, Princeton, N.J.

Bakst, M.R. 2009. Oviducal sperm and fertilisation in poultry. *Avian Biol Res* **2**: 1–5.

Bates, D., Maechler, M. & Bolker, B. (2012) lme4: linear mixed-effects models using S4 classes.
R package version 0.999999-0.http://CRAN.R-project.org/package=lme4.

Birkhead, T.R. 2010. How stupid not to have thought of that: post-copulatory sexual selection. *Journal of Zoology* **281**: 78–93.

Birkhead, T.R. 1998. Sperm competition in birds. *Rev Reprod* **3**: 123–129.

Birkhead, T.R. & Fletcher, F. 1995. Male phenotype and ejaculate quality in the zebra finch *Taeniopygia guttata*. *Proceedings Of The Royal Society B-Biological Sciences* **262**: 329–334.

Birkhead, T.R. & Fletcher, F. 1994a. Numbers of spermatozoa attached to and penetrating perivitelline layers of Japanese quail eggs. *The Auk* **111**: 997–1000.

Birkhead, T.R. & Fletcher, F. 1994b. Sperm storage and the release of sperm from the sperm storage tubules in Japanese quail *Coturnix japonica*. *Ibis* **136**: 101–104.

Birkhead, T.R. & Møller, A.P. 1998. *Sperm competition and sexual selection*. Academic Press, San Diego.

Birkhead, T.R. & Pizzari, T. 2002. Postcopulatory sexual selection. *Nature Reviews Genetics* **3**: 262–273.

Birkhead, T.R., Martinez, J.G., Burke, T. & Froman, D.P. 1999. Sperm mobility determines the outcome of sperm competition in the domestic fowl. *Proceedings Of The Royal Society B-Biological Sciences* **266**: 1759–1764.

Biswas, A., Ranganatha, O., Mohan, J. & Sastry, K. 2007. Relationship of cloacal gland with testes, testosterone and fertility in different lines of male Japanese quail. *Animal Reproduction Science* **97**: 94–102.

Biswas, A., Ranganatha, O.S. & Mohan, J. 2010. The effect of different foam concentrations on sperm motility in Japanese quail. *Veterinary Medicine International* **2010**: 1–4.

Boutin-Ganache, I., Raposo, M., Raymond, M. & Deschepper, C.F. 2001. M13-tailed primers improve the readability and usability of microsatellite analyses performed with two different allele-sizing methods. *Biotechniques* **31**: 24–6, 28.

Brillard, J.P. & Antoine, H. 1990. Storage of sperm in the uterovaginal junction and its incidence on the numbers of spermatozoa present in the perivitelline layer of hens' eggs. *British Poultry Science* **31**: 635–644.

Cameron, E., Day, T. & Rowe, L. 2007. Sperm competition and the evolution of ejaculate composition. *The American Naturalist* **169**: E158–E172.

Chapman, T. 2001. Seminal fluid-mediated fitness traits in Drosophila. *Heredity* **87**: 511–521.

Chapman, T. & Davies, S. 2004. Functions and analysis of the seminal fluid proteins of male *Drosophila melanogaste*r fruit flies. *Peptides* **25**: 1477–1490.

Chapman, T., Neubaum, D.M., Wolfner, M.F. & Partridge, L. 2000. The role of male accessory gland protein Acp36DE in sperm competition in *Drosophila melanogaster*. *Proceedings Of The Royal Society B-Biological Sciences* **267**: 1097–1105.

Cheng, K.M., Hickman, A.R. & Nichols, C.R. 1989a. Role of the proctodeal gland foam of male Japanese quail in natural copulations. *Auk* **106**: 279–285.

Cheng, K.M., McIntyre, R.F. & Hickman, A.R. 1989b. Proctodeal gland foam enhances competitive fertilization in domestic Japanese quail. *Auk* **106**: 286–291.

Clark, A.G., Aguade, M., Prout, T., Harshman, L.G. & Langley, C.H. 1995. Variation in sperm displacement and its association with accessory gland protein loci in *Drosophila melanogaster*. *Genetics* **139**: 189–201.

Clulow, J. & Jones, R. 1982. Production, transport, maturation, storage and survival of spermatozoa in the male Japanese quail, *Coturnix coturnix*. *Journal of Reproduction and Fertility* **64**: 259–266. Soc Reprod Fertility.

Coil, W.H. & Wetherbee, D.K. 1959. Observations on the cloacal gland of the Eurasian quail *Coturnix coturnix*. *The Ohio Journal of Science* **59**: 268–270.

Cornwallis, C.K. & O'Connor, E.A. 2009. Sperm: seminal fluid interactions and the adjustment of sperm quality in relation to female attractiveness. *Proceedings Of The*

Royal Society B-Biological Sciences **276**: 3467–3475.

Dean, M.D., Findlay, G.D., Hoopman, M.R., Wu, C.C., MacCoss, M.J., Swanson, W.J., *et al.* 2011. Identification of ejaculated proteins in the house mouse (*Mus domesticus*) via isotopic labeling. *BMC Genomics* **12**: 306.

Delville, Y., Sulon, J. & Balthazart, J. 1986. Diurnal variations of sexual receptivity in the female Japanese quail (*Coturnix coturnix japonica*). *Hormones and Behavior* **20**: 13–33.

Fedorka, K.M., Winterhalter, W.E. & Ware, B. 2010. Perceived sperm competition intensity influences seminal fluid protein production prior to courtship and mating. *Evolution* **65**: 584–590.

Froman, D.P., Pizzari, T., Feltmann, A.J., Castillo-Juarez, H. & Birkhead, T.R. 2002. Sperm mobility: mechanisms of fertilizing efficiency, genetic variation and phenotypic relationship with male status in the domestic fowl, *Gallus gallus domesticus*. *Proceedings Of The Royal Society B-Biological Sciences* **269**: 607–612.

Gilchrist, A. & Partridge, L. 1995. Male identity and sperm displacement in *Drosophila melanogaster*. *J Insect Physiol* **41**: 1087–1092.

Harrell, FE, Jr., (2012) Hmisc: Harrell Miscellaneous. R package version 3.10-1, URL: http://biostat.mc.vanderbilt.edu/Hmisc

Harshman, L.G. & Prout, T. 1994. Sperm displacement without sperm transfer in Drosophila melanogaster. *Evolution* **48**: 758–766.

Hodgson, D.J. & Hosken, D.J. 2006. Sperm competition promotes the exploitation of rival ejaculates. *J Theor Biol* **243**: 230–234.

Ikeda, K. & Taji, K. 1954. On the foamy ejaculate of Japanese quail, *Coturnix coturnix japonica*. *Scientific reports of the Matsuyama agricultural college* **13**: 1–4.

Kayang, B., Inoue-Murayama, M., Hoshi, T., Matsuo, K., Takahashi, H., Minezawa, M., *et al.* 2002. Microsatellite loci in Japanese quail and cross-species amplification in chicken and guinea fowl. *Genet Sel Evol* **34**: 233.

Kayang, B., Inoue-Murayama, M., Nomura, A., Kimura, K., Takahashi, H., Kimura, K., *et al.* 2000. Fifty microsatellite markers for Japanese quail. *Journal Of Heredity* **91**: 502–505.

Kilgallon, S.J. & Simmons, L.W. 2005. Image content influences men's semen quality. *Biology Letters* **1**: 253–255.

King, A.S. 1981. Cloaca. In: *Form and function in birds* (A. S. King & J. McLelland, eds), pp. 63–105. Academic Press, New York.

Klemm, R., Knight, C.E. & Stein, S. 1973. Gross and microscopic morphology of glandula-

proctodealis (foam gland) of *Coturnix c. japonica* (aves). *J Morphol* **141**: 171–184.

Korn, N., Thurston, R., Pooser, B. & Scott, T. 2000. Ultrastructure of spermatozoa from Japanese quail. *Poultry Sci* **79**: 407–414.

Lake, P. 1981. Male genital organs. In: *Form and function in birds* (A. S. King & J. McLelland, eds), pp. 1–61. Academic Press, London.

Lemaitre, J.F., Ramm, S.A., Hurst, J.L. & Stockley, P. 2011. Social cues of sperm competition influence accessory reproductive gland size in a promiscuous mammal. *Proceedings Of The Royal Society B-Biological Sciences* **278**: 1171–1176.

Locatello, L., Poli. F., Rasotto. M.B. 2013. Tactic-specific differences in seminal fluid influence sperm performance. *Proceedings Of The Royal Society B-Biological Sciences* **280**: 20122891.

Marks, H. & Lepore, P. 1965a. A procedure for artificial insemination of Japanese quail. *Poultry Sci* **44**: 1001–&.

Marks, H. & Lepore, P. 1965b. A simple intravaginal technique for artificial insemination of Japanese quail. *Poultry Sci* **44**: 1396–&.

Marshall, T.C., Slate, J., Kruuk, L.E. & Pemberton, J.M. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology* **7**: 639–655.

McFarquhar, A.M. & Lake, P.E. 1964. Artificial insemination in quail and the production of chicken-quail hybrids. *Journal of Reproduction and Fertility* **8**: 261–263. Soc Reprod Fertility.

Mohan, J., Moudgal, R., Sastry, K., Tyagi, J. & Singh, R. 2002. Effects of hemicastration and castration on foam production and its relationship with fertility in male Japanese quail. *Theriogenology* **58**: 29–39.

Møller, A.P. 1991. Sperm competition, sperm depletion, paternal care, and relative testis size in birds. *The American Naturalist*.

Neubaum, D.M. & Wolfner, M.F. 1999. Mated *Drosophila melanogaster* females require a seminal fluid protein, Acp36DE, to store sperm efficiently. *Genetics* **153**: 845–857.

Nichols, C.R. 1991. A comparison of the reproductive and behavioural differences of feral and domestic Japanese quail. University of British Columbia, masters thesis.

Ogasawara, F.X. & Huang, R. 1963. A modified method of artificial insemination in the production of chicken-quail hybrids. *Poultry Sci* **42**: 1386–1392. Poultry Science Association.

Parker, G.A. 1970. Sperm competition and its evolutionary consequences in the insects.

Biological Reviews **45**: 525–567.

Parker, G.A. 1998. Sperm competition and the evolution of ejaculates: towards a theory base. In: *Sperm competition and sexual selection* (T. R. Birkhead & A. P. Møller, eds), pp. 3–49. Academic Press, San Diego.

Parker, G.A. & Pizzari, T. 2010. Sperm competition and ejaculate economics. *Biol Rev* **85**: 897–934.

Peng, J., Chen, S., Büsser, S., Liu, H., Honegger, T. & Kubli, E. 2005. Gradual release of sperm bound sex-peptide controls female postmating behavior in *Drosophila*. *Curr Biol* **15**: 207–213.

Perry, J.C. & Rowe, L. 2010. Condition-dependent ejaculate size and composition in a ladybird beetle. *Proceedings Of The Royal Society B-Biological Sciences* **277**: 3639–3647.

Pitnick, S., Wolfner, M.F. & Suarez, S.S. 2008. Ejaculate–female and sperm–female interactions. In: *Sperm Biology: An evolutionary perspective* (T. R. Birkhead, D. J. Hosken, & S. Pitnick, eds), pp. 247-304. Academic Press, London.

Pizzari, T., Cornwallis, C.K., Froman, D.P. 2007. Social competitiveness associated with rapid fluctuations in sperm quality in male fowl. *Proceedings Of The Royal Society B-Biological Sciences* **274**: 853-860.

Pizzari, T. & Parker, G.A. 2008. Sperm competition and sperm phenotype. In: *Sperm Biology: An evolutionary perspective* (T. R. Birkhead, D. J. Hosken, & S. Pitnick, eds), pp. 207–245. Academic Press, London.

Pizzari, T., Worley, K., Burke, T., Froman, D.P. 2008. Sperm competition dynamics: ejaculate fertilising efficiency changes differentially with time. *BMC Evolutionary Biology* **8**: 33

Poiani, A. 2006. Complexity of seminal fluid: a review. *Behav Ecol Sociobiol* **60**: 289–310.

Prout, T. & Clark, A.G. 2000. Seminal fluid causes temporarily reduced egg hatch in previously mated females. *Proceedings Of The Royal Society B-Biological Sciences* **267**: 201–203.

Ram, K.R. & Wolfner, M. 2007. Seminal influences: *Drosophila* Acps and the molecular interplay between males and females during reproduction. *Integr Comp Biol* **47**: 427–445.

Ram, K.R. & Wolfner, M.F. 2009. A network of interactions among seminal proteins underlies the long-term postmating response in *Drosophila*. *Proceedings of the National Academy of Sciences* **106**: 15384–15389.

Ram, K.R. & Wolfner, M.F. 2005. Sustained post-mating response in *Drosophila melanogaster* requires multiple seminal fluid proteins. *Plos Genetics* **3**: 2428–2438.

Ramm, S.A., Parker, G.A. & Stockley, P. 2005. Sperm competition and the evolution of male reproductive anatomy in rodents. *Proceedings Of The Royal Society B-Biological Sciences* **272**: 949–955.

R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Seiwert, C. & Adkins-Regan, E. 1998. The foam production system of the male Japanese quail: Characterization of structure and function. *Brain Behav Evol* **52**: 61–80.

Simmons, L. & Fitzpatrick, J. 2012. Sperm wars and the evolution of male fertility. *Reproduction* **144**: 519–534.

Simmons, L.W. 2001. *Sperm Competition and Its Evolutionary Consequences in the Insects*. Princeton University Press, Princeton, NJ.

Singh, R.P., Sastry, K.V.H., Pandey, N.K., Shit, N., Singh, R. & Mohan, J. 2011a. Sperm competition in Japanese quail (*Coturnix japonica*): Last male precedence is declined in two successive matings with two different males. *Indian Journal of Poultry Science* **46**: 130–131.

Singh, R.P., Sastry, K.V.H., Shit, N., Pandey, N.K., Singh, K.B., Mohan, J., *et al.* 2011b. Cloacal gland foam enhances motility and disaggregation of spermatozoa in Japanese quail (*Coturnix japonica*). *Theriogenology* **75**: 563–569.

Singh, R.P., Sastry, K.V.H., Pandey, N.K., Singh, K.B., Malecki, I.A., Farooq, U., *et al.* 2012. The role of the male cloacal gland in reproductive success in Japanese quail (*Coturnix japonica*). *Reprod. Fertil. Dev.* **24**: 405.

Sirot, L.K., Laflamme, B.A., Sitnik, J.L., Rubinstein, C.D., Avila, F.W., Chow, C.Y., *et al.* 2009. Molecular social interactions: *Drosophila melanogaster* seminal fluid proteins as a case study. *Adv. Genet.* **68**: 23–56.

Sirot, L.K., Wolfner, M.F. & Wigby, S. 2011. Protein-specific manipulation of ejaculate composition in response to female mating status in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences* **108**: 9922–9926.

Sittmann, K. & Abplanalp, H. 1965. Duration and recovery of fertility in Japanese quail (*Coturnix coturnix japonica*). *British Poultry Science* **6**: 245–250.

Smith, C.C. & Ryan, M.J. 2011. Tactic-dependent plasticity in ejaculate traits in the swordtail *Xiphophorus nigrensis*. *Biology Letters* **7**: 733–735.

Snook, R. 2005. Sperm in competition: not playing by the numbers. *Trends Ecol Evol* **20**: 46–53.

Teijeiro, J.R., Puigcerver, M. & Gallego, S. 2003. Pair bonding and multiple paternity in the polygamous common quail *Coturnix coturnix*. *Ethology* **109**: 291–302.

Wang, Z., Gerstein, M. & Snyder, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10**: 57–63.

Wentworth, B.C. & Mellen, W.J. 1963. Egg production and fertility following various methods of insemination in Japanese quail (Coturnix coturnix japonica). *Journal of Reproduction and Fertility* **6**: 215–220. Soc Reprod Fertility.

Wickham, H. 2009. *ggplot2: elegant graphics for data analysis*. Springer Publishing Company, Incorporated.

Wigby, S., Sirot, L.K., Linklater, J.R., Buehner, N., Calboli, F.C.F., Bretman, A., *et al.* 2009. Seminal fluid protein allocation and male reproductive success. *Curr Biol* **19**: 751–757.

Wilson, W.O. & Huang, R.H. 1962. A comparison of the time of ovipositing for Coturnix and chicken. *Poultry Sci* **41**: 1843–1845. Poultry Science Association.

Wilson, W.O., Abbott, U.K. & Abplanalp, H. 1961. Evaluation of Coturnix (Japanese quail) as pilot animal for poultry. *Poultry Sci* **40**: 651–657. Poultry Science Association.

Wolfner, M.F. 1997. Tokens of love: functions and regulation of Drosophila male accessory gland products. *Insect Biochemistry and Molecular Biology* **27**: 179–192.

CHAPTER 2

A COMPARISON OF NEXT-GENERATION SEQUENCING APPROACHES FOR TRANSCRIPTOME ASSEMBLY AND UTILITY FOR RNA-SEQ IN A NON-MODEL SPECIES

Abstract

Next-generation sequencing affords researchers working on non-model organisms the opportunity to sequence transcriptomes without reference genomes. *De novo* assembled transcriptomes, in combination with RNA-Seq, are powerful tools to explore ecological and evolutionary hypotheses at both a gene sequence and expression level. Investigators first must choose which of the two juggernaut sequencing platforms, 454 or Illumina, will provide data most suitable for their experimental goals. We present a case study exploring the utility of 454 and Illumina sequences for *de novo* transcriptome assembly and downstream RNA-Seq applications in a reproductive gland from a non-model bird species. Four transcriptomes composed of either pure 454 or Illumina reads or mixtures of read types were assembled and evaluated. Hybrid assemblies generally performed best for *de novo* transcriptome characterization in terms of contig length, transcriptome coverage, gene annotation, and complete assembly of gene transcripts. Improvements over the Illumina assembly were negligible. The Illumina assembly provided the best scaffold to map an independent set of RNA-Seq data as ~ 83.5% of reads mapped to a single contig in the transcriptome. Scaffolds constructed solely from 454 data may impose problems for RNA-Seq as ours revealed a high number of indels and many ambiguously mapped reads. Illumina and a combination of Illumina and 454 data produce high quality transcriptomes appropriate for RNA-Seq gene expression analyses without a reference genome. In our non-model avian species, sequencing with Illumina alone allowed for transcriptome assembly and quantification of expression profiles in a single sequencing step.

Introduction

Until recently, evolutionary and population-genomic research was restricted to the small number of taxa considered model organisms. Modern next-generation sequencing technologies offer the opportunity to generate massive (and increasing) amounts of sequence data easily and affordably. Today, the potential for large-scale genomic investigations exists for virtually any study system (Ellegren, 2008; Hudson, 2008; Ekblom & Galindo, 2010; Wheat, 2010). One popular and fruitful approach adopted by the non-model research community is shotgun-sequencing of transcriptomes (Wheat, 2010; Ekblom & Galindo, 2010). With the advent of deep, parallel sequencing of RNA, termed RNA-Seq, researchers can additionally quantify expression variation in a high-throughput and cost-effective manner (Wang *et al.*, 2009). Given options in terms of sequencing platform and bioinformatics workflow, a pressing question is what is the optimal strategy to harness both the static (sequence-level) and dynamic (expression-level) nature of transcriptomes of non-model species on limited budgets.

In the last five years, investigators predominantly utilized long sequencing reads generated by 454 GS-FLX (Roche Diagnostics Corporation; hereafter "454") sequencing platform to facilitate *de novo* transcriptome assembly (Vera *et al.*, 2008; Meyer *et al.*, 2009; Schwartz *et al.*, 2010; Ekblom & Galindo, 2010; Wheat, 2010; Reading *et al.*, 2012; Ekblom *et al.*, 2012). While 454 is appropriate for assembly, the millions of short reads produced by Illumina© (Illumina, Inc.) are preferred for RNA-Seq as detection of differential expression is sensitive to sequencing depth (Wang *et al.*, 2009; Ekblom & Galindo, 2010; Tarazona *et al.*, 2011; Wang *et al.*, 2011). One approach to RNA-Seq has been to map Illumina short reads onto scaffolds constructed from longer 454 reads (*e.g.*, Su *et al.*, 2011;

Jensen *et al.*, 2012). With increasing read lengths produced by Illumina HiSeq™ technology (currently 100 or 150 bp, depending on the run), studies assembling *de novo* transcriptomes directly from Illumina data are emerging (*e.g.*, Crawford *et al.*, 2010; Chen *et al.*, 2010; Birzele *et al.*, 2010; Feldmeyer *et al.*, 2011; Xia *et al.*, 2011; Etebari *et al.*, 2011; Van Belleghem *et al.*, 2012a). This approach is attractive, as data for transcriptome characterization and quantification are collected at once (Chen *et al.*, 2010; Crawford *et al.*, 2010; Birzele *et al.*, 2010; Etebari *et al.*, 2011). Recent work comparing technologies using both real and simulated data suggest that hybrid assemblies combining 454 and Illumina reads yield the highest quality transcriptomes (Wall *et al.*, 2009; Hornett & Wheat, 2012; Cahais *et al.*, 2012). However, collecting both types of data may be cost-prohibitive. Here, we sequence a transcriptome of a non-model organism with both 454 and Illumina technologies, perform *de novo* assembly with each data type separately or in some combination, and compare the various transcriptomes in terms of quality and utility for RNA-Seq. Our objective was to model the approach that a non-model researcher would take and considered cost as a limiting factor. Therefore, we sequenced our transcriptome with both 454 and Illumina technology for approximately the same cost (~$5000, Table S2.1) and assembled each type of data with assemblers known to work well with the two technologies.

Our transcriptome data derives from a reproductive tissue of a non-model species. Male Japanese quail (*Coturnix japonica*) possess a well-developed foam gland that secretes a viscous protein complex that is whipped into a stiff foam by contractions of the cloacal sphincter muscle (Klemm *et al.*, 1973; Seiwert & Adkins-Regan, 1998; Finseth *et al.*, 2013a). During copulation, a male introduces semen, along with a large quantity of foam, to a

female's reproductive tract (Coil & Wetherbee, 1959). This trait is of interest to evolutionary biologists because it is an example of a novel trait (Klemm *et al.*, 1973; Fujihara, 1992; Finseth *et al.*, 2013b), it is likely involved in sexual selection (Finseth *et al.*, 2013c, Cheng *et al.*, 1989b), and it encodes proteins transferred from males to females during mating, which have been shown to be key mediators of reproductive success (Poiani, 2006). Foam enhances male reproductive fitness in a variety of ways including mediating the outcome of sperm competition and improving several aspects of fertility and sperm performance (Cheng *et al.*, 1989a,b; Adkins-Regan, 1999; Singh *et al.*, 2012; Singh *et al.*, 2011; Finseth *et al.*, 2013c). Characterizing the foam gland transcriptome can offer insight into how sexual selection shapes the genome, as well as the nature of genetic changes responsible for novel traits.

We sequenced the foam gland with both 454 and Illumina technologies and assembled transcriptomes following four different schemes that have been applied to species without genomic resources (Wall *et al.*, 2009; Ekblom & Galindo, 2010; Milano *et al.*, 2011; Su *et al.*, 2011; Cahais *et al.*, 2012; Hornett & Wheat, 2012). We initially evaluated the transcriptome assemblies with standard metrics (*e.g.*, N50, median contig length, etc.). Several features of transcriptomes (*e.g.* repetitive sequences, unequal coverage, alternative splicing, and heterozygosity) present challenges for *de novo* assembly (Sammeth, 2009; Surget-Groba & Montoya-Burgos, 2010; Vijay *et al.*, 2012). Resulting transcriptomes may retain errors that would not be captured by standard metrics (*e.g.*, sequencing errors, insertions/deletions, misassembled paralogs, chimeras, and/or partial transcripts), but could introduce bias for downstream applications like RNA-Seq (Vijay *et al.*, 2012; Cahais *et al.*, 2012). Well-annotated genomes can help gauge the frequency of such errors. While

Japanese quail do not have a sequenced genome available, functional annotation using a related species' genome as a proxy reference is robust for species pairs diverged less than 100 million years (Hornett & Wheat, 2012). Japanese quail diverged ~ 34 million years ago from the chicken (*Gallus gallus*) and exhibit conserved synteny and chromosomal structure with the chicken genome (Kayang *et al.*, 2006; Sasazaki *et al.*, 2006a; Sasazaki *et al.*, 2006b). To estimate the incidence of confounding assembly errors, we annotated our quail transcriptomes with the chicken genome and assessed how well assembled contigs reproduced orthologous genes. Finally, we evaluated each transcriptome's utility for RNA-Seq by mapping short read data from an independent sampling of foam glands to the assemblies and comparing the alignments.

Methods

Sample collection and RNA extraction

Japanese quail were lab-reared and housed on a 16:8 light:dark cycle to simulate breeding conditions. All study males were sexually mature, had phenotypically normal foam glands, and produced normal foam complements. A foam gland from a Japanese quail male approximately one year old was used to generate the 454 data. Foam glands from six Japanese quail males (two were one-year old and four were five months old), were used to generate the Illumina data for transcriptome assembly. For the independent RNA-seq assessment, we sampled foam glands from six different Japanese quail males on winter light conditions (8:16 light:dark cycle, with lights on at 8:00) with testosterone replacement. Testosterone-replaced males have phenotypically normal foam glands, produce foam, and were the same ages as the individuals used to build the transcriptomes

(Adkins, 1977; Schumacher & Balthazart, 1983). After euthanizing with CO_2, we immediately dissected out foam glands and froze samples on liquid nitrogen. We extracted RNA with the Agencourt® RNAdvance™ Tissue Kit (Beckman Coulter) following the manufacturer's instructions with the exception that we performed half-reactions. RNA quality and concentration was assessed by agarose gel electrophoresis and NanoDrop™ spectrophotometry. We checked for RNA purity and integrity using an Agilent 2100 BioAnalyzer.

Library construction

454

In March 2010, we isolated mRNA from approximately one μg total RNA, synthesized first-strand cDNA and generated ds cDNA following the manufacturer's instructions for the SMART™ Polymerase Chain Reaction (PCR) cDNA Synthesis Kit (Clontech Laboratories, Inc.), with the exception that we used SuperScript™ III Reverse Transcriptase (Invitrogen) as the reverse transcriptase and made adjustments accordingly. To generate ds cDNA, we followed the suggested protocol except used Platinum® *Taq* DNA Polymerase (Life Technologies). We performed long distance PCR, confirmed successful amplification via agarose gel electrophoresis, and cleaned the PCR products with the QIAquick PCR Purification Kit (Qiagen). We partially normalized our ds cDNA pool in an effort to retain a crude signature of abundance while increasing the representation of rare transcripts. For partial normalization, amplified cDNA was subjected to hybridization and double-stranded nuclease (DSN) digestion following instructions from the TRIMMER cDNA Normalization Kit (Evrogen) except using only 1/8 and 1/16 concentration of DSN. Size

selection was performed with the QIAquick Gel Extraction Kit (Qiagen) according to manufacturer's instructions. We enzymatically fragmented the dsDNA with NEBNext™ dsDNA Fragmentase (New England BioLabs®, Inc.), end polished using T4 polymerase (New England BioLabs®, Inc.), phosphorylated 5' ends with T4 kinase (New England BioLabs®, Inc.), added an adenine to 3' ends with NEB Taq (New England BioLabs®, Inc.), and ligated Multiplex Identifier (MID) Adaptor #1 for GS FLX Titanium chemistry (Roche/454 Life Sciences) to ds cDNA using T4 ligase (New England BioLabs®, Inc.). Throughout the normalization, end polishing, and ligation procedure, the ds cDNA was cleaned with the QIAquick PCR Purification Kit (Qiagen). Resulting concentration was determined using NanoDrop™ spectrophotometry. Cornell University's Genomics Facility at the Institute of Biotechnology performed ½ plate of 454 GS FLX sequencing with Titanium chemistry on the resulting library (Roche/454 Life Sciences) in April 2010.

Illumina

In January 2012, six Illumina libraries for the transcriptome assembly were prepared from approximately 1.2 µg total RNA using the TruSeq™ RNA Sample Preparation Kit (Illumina®) following the manufacturer's instructions. We also prepared six samples from testosterone-replaced males for the independent RNA-seq evaluations. All twelve samples were tagged with a unique adapter index, pooled, and single-end sequenced on one lane of an Illumina HiSeq™ 2000, with a target read length of 100 bp. Sequencing was performed by Cornell University's Genomics Facility at the Institute of Biotechnology in April 2012.

De novo *assembly*

Our objective was to construct four high quality assemblies economically, thereby mimicking the approach adopted by non-model researchers. Many studies prior have compared different transcriptome assemblers and there is some consensus regarding which assemblers perform optimally on 454 and Illumina sequences (Kumar & Blaxter, 2010; Grabherr *et al.*, 2011; Zheng *et al.*, 2011; Feldmeyer *et al.*, 2011; Rawat *et al.*, 2012). Thus, we chose to use freely available assemblers previously shown to be appropriate for each data type. For 454 data, combining output from multiple assemblers gives the best end product (Kumar & Blaxter, 2010). Therefore, we used the assembly pipeline iAssembler, which performs iterative assemblies with MIRA (4 cycles) and CAP3 (1 cycle), followed by automated error detection and correction (Zheng *et al.*, 2011). For Illumina data we chose to use the Trinity assembler, which has effectively reconstructed many transcriptomes from Illumina data (Grabherr *et al.*, 2011; Van Belleghem *et al.*, 2012).

All sequence manipulations and assemblies (save the Trinity assembly) were performed on a Linux Dell Precision T3500n with 4 cores, 24 GB RAM, and 4 TB HDD housed at Cornell University's Bioinformatics Facility at the Institute of Biotechnology. The Trinity assembly was executed on a Linux Dell PowerEdge R710 with 16 cores, 64 GB RAM, and 1 TB HDD also at the same place.

454

Initial quality filtering of reads was performed by the Cornell University's Genomics Facility at the Institute of Biotechnology. SeqClean (http://compbio.dfci.harvard.edu/tgi/software/) was used to trim low complexity

sequences and short sequences (< 90 bp). MID-1 and SMART adaptors were trimmed using both SeqClean and NextGENE® (Softgenetics®). We assembled the reads into 88,840 contigs using two rounds of iAssembler with default parameters (Zheng *et al.*, 2011). The output from the first round of iAssembler was used for the second round. CD-HIT EST collapsed contigs into unigenes, with default parameters except we set sequence identity to 95% and compared both strands (Huang *et al.*, 2010). We removed any reads that did not join at least one other read to form a contig in iAssembler (*i.e.,* a singleton) or were not identified as part of a cluster using CD-HIT EST, resulting in 27,362 contigs. To remove redundancy, we identified redundant contigs using BLAT with our library as both query and subject (*i.e.*, self-BLAT). BLAT identifies DNA sequences with >95% similarity. The smaller member of a significant alignment was removed if 1) no more than 100 bp of the smaller member remained unaligned and/or 2) the alignment covered the length of at least 85% of the smaller pair (3,798 contigs removed).

Illumina

Initial quality filtering and barcode removal were performed by Cornell University's Genomics Facility at the Institute of Biotechnology. We used fastq-mcf (http://code.google.com/p/ea-utils/wiki/FastqMcf) to remove Illumina adaptors, trim low-quality terminal ends, discard short sequences, and filter reads with phred scores < 20. We merged the six libraries into a single file and assembled a transcriptome using the Trinity pipeline with default parameters (Grabherr *et al.*, 2011). The resulting assembly had 39,694 contigs. We intended to remove redundant entries with > 95% similarity as with the 454 assembly. However, we found a very small number of contigs that met our

criteria (106, <0.003% of the assembly) and therefore retained the entire contig set generated by Trinity.

Corrected 454

We used the Illumina data to correct errors with our 454 transcriptome using the Nesoni pipeline, version 0.85 (http://www.bioinformatics.net.au/software.nesoni.shtml). Nesoni inputs a set of reference sequences and a set of short reads, and utilizes the SHRiMP short read mapper to align reads to the reference (Rumble *et al.*, 2009). Nesoni then identifies positions where disparity exists between the majority of reads and the reference. It corrects errors based on a set of criteria, creating a corrected consensus sequence set. We input the 454 transcriptome as reference, the Illumina data as reads, and created consensus sequences using default parameters. Only those transcripts with at least one aligned read were retained. Seqclean with default parameters was used to trim N's from the consensus sequence (http://compbio.dfci.harvard.edu/tgi/software/). The cleaned consensus sequence set represents the "corrected 454" assembly.

Hybrid

Contigs from the Corrected 454 and Illumina assemblies were merged and assembled into a hybrid transcriptome assembly using iAssembler with default parameters.

Transcriptome evaluation

Unless specified, analyses were conducted in R version 2.15.1 and RStudio version 0.96.330.

Standard metrics

For each assembly, we calculated standard metrics of quality including number of contigs, average contig length, median contig length, N50 (median contig size, weighted by length), the distribution of contig lengths, and summed contig length (Kumar & Blaxter, 2010; Hornett & Wheat, 2012). We downloaded all chicken coding sequences from Ensembl version 68 (*G. gallus* assembly: WASHUC2) via the BioMart tool (http://www.biomart.org/) and calculated the same standard metrics for comparison purposes. Prior to computation of basic metrics, we removed contigs less than 200 bp in each dataset, as Trinity assemblies do not report contigs less than 200 bp. We were also interested in how each assembly performed in terms of predicting open reading frames. We identified open reading frames with OrfPredictor (Min *et al.*, 2005). For each assembly, we computed the frequency of contigs with no open reading frames and the distribution of the lengths of open reading frames.

Ortholog comparisons

We used data from chickens to identify orthologs. All chicken protein sequences from Ensembl version 68 (*G.* gallus assembly: WASHUC2) were downloaded via the BioMart tool (http://www.biomart.org/). We filtered the protein set to remove redundant entries (*i.e.*, duplicates, alternative splice variants) by self-BLAST following Hornett and Wheat (2012). In brief, for any pairwise BLASTp hit with an e-value $\leq 1 * 10^{-6}$, > 90%

similarity, and > 33 amino acids in length, we removed the shorter of the two proteins. All BLAST steps were performed in parallel via Cornell University's Computational Biology Application Suite for High Performance Computing (biohpc.org). The reciprocal best blast method was used to determine orthologs (Tatusov, 1997; Bork & Koonin, 1998; Koonin, 2005). In short, we BLASTed the four assemblies against the filtered Ensembl chicken protein sequences, using the appropriate BLAST packages with a cutoff e-value of $1 * 10^{-6}$ and vice versa. Orthologs were called when the top chicken hit from the quail to chicken BLAST returned the original quail query in the chicken to quail BLAST. We report the number of orthologs identified for each transcriptome assembly. We also present the distribution of orthologs identified across the various assemblies in a Venn diagram made using the VennDiagram package in R (Chen & Boutros, 2011).

For each contig from the various transcriptome assemblies, we computed the "ortholog hit ratio" as described by O'Neil *et al*. (2010) and implemented by Van Belleghem *et al*. (2012). This ratio represents the length of a putative coding region of a contig divided by the length of the coding region of its orthologous transcript. The hit region of the best BLASTx result between a contig and its ortholog was used as a conservative estimate of the "putative coding region" of a contig. Lengths are in amino acids. An ortholog completely represented by a contig would have a ratio of "1". Ratios less than 1 indicate instances where contigs only partially covered orthologs, while ratios greater than 1 indicate insertions in contigs.

Independent RNA-Seq assessment

We derived independent Illumina RNA-Seq data from foam glands of six different foam-producing Japanese quail males to evaluate the utility of our various assemblies for gene expression analyses. The RNA-Seq data were merged into a single file and aligned using the Burrow-Wheeler transform as implemented in BWA (Li & Durbin, 2009). We employed the Nesoni pipeline (http://vicbioinformatics.com/nesoni.shtml) to generate statistics about the quality of the alignments to the various transcriptome assemblies including the number of mapped and unmapped reads, and the number of insertions and deletions per 100,000 bp identified between each assembly and the majority of the RNA-Seq data. We calculated the number of uniquely mapped reads with samtools (http://samtools.sourceforge.net/).

Results and Discussion

Transcriptome sequencing approaches

Foam glands were sequenced with Illumina and 454 data (see Methods). To clarify, a "read" refers to the raw sequence data produced from one of the sequencing platforms and a "contig" is a sequence that results from assembling multiple reads (*i.e.*, one unit of an assembled transcriptome. The two types of data were used solo or in combination to assemble four different transcriptomes. The first assemblies were composed solely of reads from one or the other technology ("454" and "Illumina" assemblies). The third strategy attempted to address known issues with systematic errors inherent to 454 sequencing (*e.g.*, homopolymer errors; Hudson, 2008; Gilles *et al.*, 2011). For this approach, we mapped Illumina reads onto the 454 assembly, identified points of discrepancy between the 454 contig and the majority of Illumina reads, and created a

corrected consensus sequence ("Corrected 454"). Finally, we constructed a hybrid assembly ("Hybrid") by merging contigs made by the Corrected 454 and Illumina assemblies. We chose to merge contigs, rather than assemble from raw reads, because recent work in non-model systems suggests that this method performs better than a merge-reads hybrid approach in terms of contig length, total transcriptome coverage, and number of genes identified (Cahais *et al.*, 2012). We report raw data for the sequencing runs in (Table S2.1). In an effort to mimic the approach a non-model researcher we used, we chose assemblers known to be appropriate for each sequencing type (see Methods).

Standard transcriptome quality assessment

High quality assemblies possess near full-length contigs representing most of the actual transcriptome. We first evaluated each transcriptome assembly and the chicken coding sequence set using a suite of standard metrics (Kumar & Blaxter, 2010; Hornett & Wheat, 2012). We assumed that estimates from the chicken coding sequence set provide a reasonable expectation, as gene length is highly conserved within eukaryotes (Xu, 2006). Table 2.1 compares metrics describing contig length and the number of contigs across assemblies. Figure 2.1 displays the distribution of contig lengths across the five datasets. Note that only contigs greater than 200 bp were included.

Standard metrics of transcriptome quality revealed that the best assemblies incorporated Illumina data either alone (Illumina) or in conjunction with 454 data (Hybrid; Table 2.1). These two assemblies resulted in longer contigs that were closer to the expected length based on the chicken sequences than either the 454 or Corrected 454 assemblies (Table 2.1). The Hybrid and Illumina assemblies generated lots of contigs that

Table 2.1: Standard metrics of transcriptome assembly (lengths and N50 in nucleotide base pairs).

Assembly	Number of contigs	N50	Median contig length	Mean contig length	Maximum contig length	Summed contig length
454	23,564	486	410	457	4,114	12,408,584
Illumina	39,694	1179	379	710	12,391	28,190,399
Corrected 454	13,211	496	394	451	2335	4,611,915
Hybrid	38,946	1227	385	728	12,391	28,197,988
Chicken coding	23,392	2136	1068	1507	26,362	32,322,198

Table 2.2: Number and frequency of contigs with or without open reading frames.

Assembly	# contigs with ORF	# contigs with no ORF	Frequency (%)
454	23,364	200	0.85
Illumina	39,466	229	0.58
Corrected 454	13,135	76	0.58
Hybrid	38,721	225	0.58
Chicken coding	17,031	2	0.01

Table 2.3: The number of deletions, and insertions per 100,000 bp identified between RNA-Seq and an assembly.

Assembly	Deletions	Insertions
454	182.87	42.08
Illumina	0.94	0.46
Corrected 454	16.57	2.26
Hybrid	1.12	0.46
Chicken coding	0.33	0.19

were long (N50, mean, longest contig) and covered a large proportion of the transcriptome (summed contig length) (Table 2.1). While the number of contigs from the 454 assembly almost replicated the number of coding sequences described from the chicken sequence set (Table 2.1), they tended to be short (N50, mean, longest contig) and had a low summed length (Table 2.1). We note that the median contigs lengths are very similar across all four transcriptomes, but that the means of the Illumina and Hybrid assemblies are much higher. This is likely because the Illumina data produced an excess of very short contigs (200 bp), but many more long contigs (Figure 2.1a, discussed below). Our findings support earlier work showing that hybrid assemblies produced the largest summed contig lengths, though previous studies were mixed regarding which technology used singly yielded the next best results (Cahais *et al.*, 2012; Hornett & Wheat, 2012). One prior study similarly found that transcriptome assemblies composed solely of Illumina reads exhibited longer contigs than those composed only of 454 reads (Hornett & Wheat, 2012), though other studies did not (Milano *et al.*, 2011; Cahais *et al.*, 2012).

The Corrected 454 assembly performed poorly across almost all basic metrics revealing short contigs representing a small portion of the expected transcriptome size (Table 2.1). This pattern arises in part because the 454 assembly constitutes an upper limit in terms of contig length and number of contigs for the Corrected 454 assembly. For example, the dramatic decrease in the number of contigs from the 454 assembly is because the Corrected 454 transcriptome is limited to consensus sequences between the 454 transcriptome and RNA-Seq data. Thus, only the subset of the 454 assembly with at least some mapped Illumina reads were retained. Improvements in the Corrected 454 versus

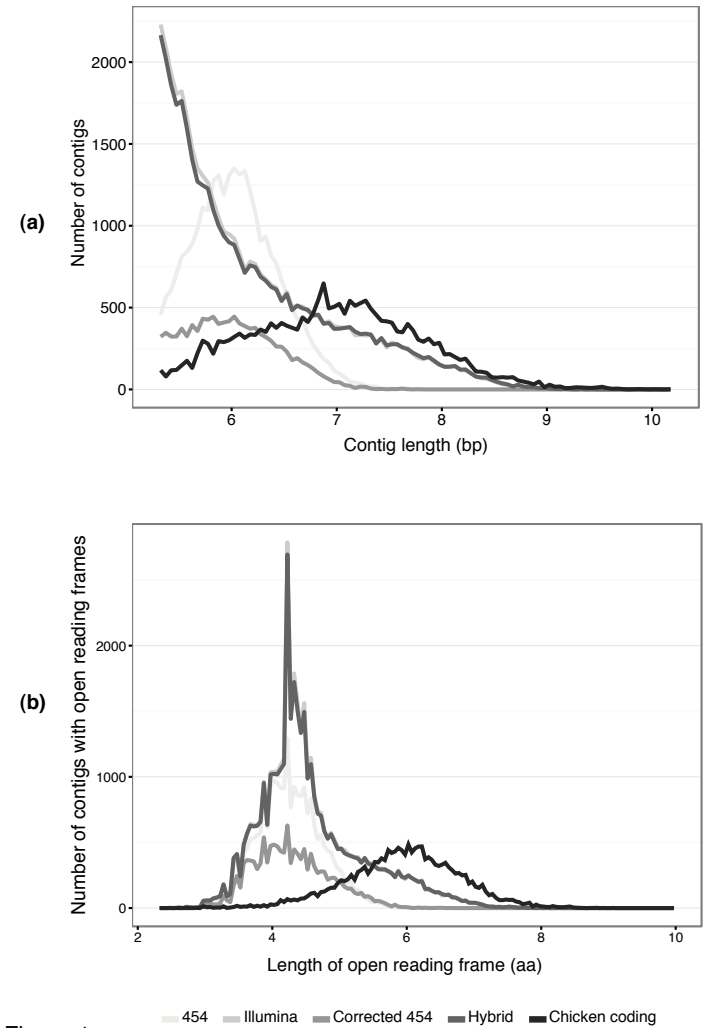

Figure 2.1: a) Histogram of contig lengths in nucleotide base pairs of each of the transcriptome assemblies and the chicken coding sequence set. **b)** Histogram of open reading frame lengths in amino acids predicted for each of the transcriptome assemblies and the chicken coding sequence set. Lengths were natural log transformed prior to plotting. Legend applies to both graphs.

the 454 transcriptomes will be seen as revised errors within the assembly and, therefore, would not be captured by standard metrics.

Figure 2.1a shows that the distribution of contig lengths are quite different across 454- and Illumina-based datasets, though similar to patterns described previously (Cahais *et al.*, 2012; Van Belleghem *et al*., 2012). Despite an excess of short contigs in the Illumina and Hybrid assemblies, the extended, broad right tails of their distributions indicate a large number of well-assembled long transcripts (Figure 2.1a). The shape of the distributions of the 454 and Corrected 454 assemblies are similar to that of chicken coding sequences. However, the right tails (*i.e.*, long contigs) of the Illumina and Hybrid distributions closely shadow the chicken sequence set, contrasting with a non-existent right tail from 454-based assemblies. The Hybrid assembly appears heavily weighted towards Illumina data, as their length distributions are virtually indistinguishable (Figure 2.1).

Some errors in transcriptome assembly (*e.g.,* homopolymer errors) can produce frameshifts, in which case downstream analyses reliant on properly called open reading frames would be difficult. Frameshifts can create premature stop codons, resulting in shorter open reading frames. We predicted open reading frames *in silico* for each transcriptome assembly and computed the frequency of contigs with no open reading frames, as well as the distribution of the lengths of open reading frames (Table 2.1, Figure 2.1b). The assemblies produced the same frequency of contigs with no reading frames (~0.58%), save the 454 assembly, which had a higher rate (~0.85%), and the chicken sequence set, which exhibited a lower rate (~.01%; Table 2.2). The lengths of open reading frames recovered distributions similar to those of the original contigs, with the Illumina and Hybrid assemblies having many more both short and long open reading frames (Figure

2.1). All of the assemblies produced relatively short open reading frames compared to the chicken coding sequence (Figure 2.1b).

The Hybrid assembly displayed the highest values across most standard metrics, but the improvement over the Illumina assembly was marginal (Table 2.1). Further, the contig length distributions of the Hybrid and Illumina assemblies are nearly identical (Figure 2.1a). Previously, approaches combining both 454 and Illumina data reveal significant improvements over either single technology as judged by basic metrics (Wall *et al.*, 2009; Cahais *et al.*, 2012). While our quantitative assessment supports this conclusion in the absolute sense, our results suggest that Illumina data alone can produce similar quality transcriptomes as hybrid strategies.

The discrepancy between our results and previous work may be due to aspects of our experimental design. Since our sequencing efforts, 454 has introduced GS FLX+ chemistry (Roche Diagnostics Corporation), which promises more reads that are longer (up to 1000 bp) than the GS FLX Titanium chemistry we used. Longer reads can improve transcriptome contiguity and reduce mis-assembly of short reads (Wall *et al.*, 2009). We chose to sequence one-half lane of Illumina and one-half plate of 454 for the transcriptome assemblies as they were approximately the same cost, but this strategy generated significantly more Illumina data (Table S2.1). The discrepancies in coverage could, therefore, explain many of the differences in transcriptome quality. However, our merge-contigs hybrid approach started with comparatively similar numbers of contigs from each transcriptome (Illumina: ~39K: Corrected 454: ~14K), so it is surprising the degree to which the Illumina data dominates the Hybrid assembly (Table 2.1, Figure 2.1). Additionally, we sequenced a single tissue that likely expresses fewer genes than an entire

organism would. Thus, assemblies generated with short Illumina reads may be appropriate for sequencing a smaller number of genes, but hybrid assemblies may exhibit improvements as the number and diversity of expressed genes increase.

Comparisons with orthologs

De novo assembled transcriptomes from non-model species rely on BLAST-based annotations to provide information about gene identity and function. We exploited the fact that quail and chicken are closely related and determined quail-chicken orthologs via reciprocal best BLAST (Bork & Koonin, 1998; Koonin, 2005). Our data suggest that assemblies derived at least in part from Illumina sequences outperform those built solely from 454 reads in terms of the number of orthologs identified and, therefore, provide researchers significantly more annotations. The Hybrid (7,323) and Illumina (7,278) assemblies revealed almost twice the number of identified orthologs than either the 454 (4,069) or Corrected 454 (4,079) datasets (Figure 2.2). Our results contrast with previous work that annotated similar (Cahais *et al.*, 2012; Hornett & Wheat, 2012) or more (Milano *et al.*, 2011) contigs in assemblies from 454 data than Illumina data. Compared to the previous studies, we either generated more Illumina and less 454 sequence data, or implemented the newer Illumina HiSeq™ 2000 sequencing technology (Table S2.1; Milano *et al.*, 2011; Cahais *et al.*, 2012). Hybrid assemblies generally performed well across studies (present study, Cahais *et al.*, 2012; Hornett & Wheat, 2012).

If a research goal is to annotate the maximum number of genes, our data suggest that combining annotations from hybrid and single-data assemblies may be the preferred method (Figure 2.2). The 454 reads contributed an additional 1,161 annotations over the

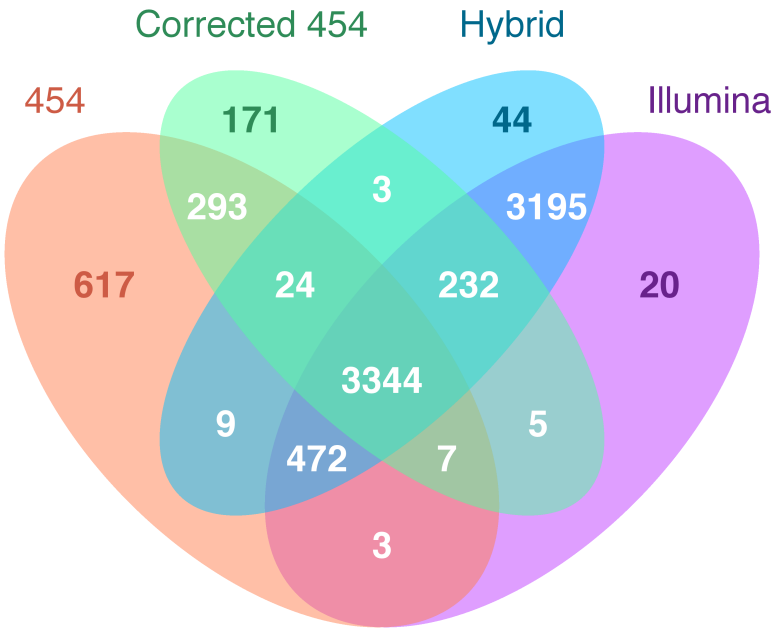

Figure 2.2: Venn diagram representing the number of orthologs identified via reciprocal best BLAST with chicken and each transcriptome assembly. Non-white numbers indicate orthologs that were unique to one assembly.

Illumina assembly, whereas the Illumina data added 3, 670 on top of the 454 transcriptome. These are significant numbers, as only 8,439 contigs were annotated in total. Thus, utilizing both types of data can substantially improve the number of gene annotations, though more so with Illumina.

The number of orthologs we characterized seems reasonable for a Galliformes transcriptome. Two transcriptomes for the Japanese quail and Northern bobwhite quail (*Colinus virgianus*) constructed from 454 data identified more orthologs than our 454 assembly (7,391 and 6,795, respectively), but roughly similar numbers as our Illumina and Hybrid assemblies (Table 2.1; Rawat *et al.*, 2010a; Rawat *et al.*, 2010b; Rawat *et al.*, 2012). However, we caution direct comparison of their results with ours as the authors implemented a reference-based, as opposed to *de novo*, assembly strategy, tissue samples were either unspecified and/or from multiple tissues (including samples that were chemically treated), and less stringent BLAST criteria were used to assign orthology.

Contigs from optimal assemblies will represent full, as opposed to partial, gene sequences. We assessed how well contigs from each assembly reproduced ortholog length by calculating the "ortholog hit ratio" (Figure 2.3; O'Neil *et al.*, 2010; Van Belleghem *et al.*, 2012). This ratio is the length of the assembled contig relative to the length of its chicken ortholog. Contigs representing fully assembled transcripts will have ortholog hit ratios close to one. Values less than one indicate contigs are fragmented, while insertions in contigs would yield values greater than one. All assemblies displayed many ratios less than one, suggesting that partial transcripts are a challenge for *de novo* assembled transcriptomes (Figure 2.3). The Illumina and Hybrid assemblies revealed many more fully assembled transcripts and a more even distribution of ortholog hit ratios than either the

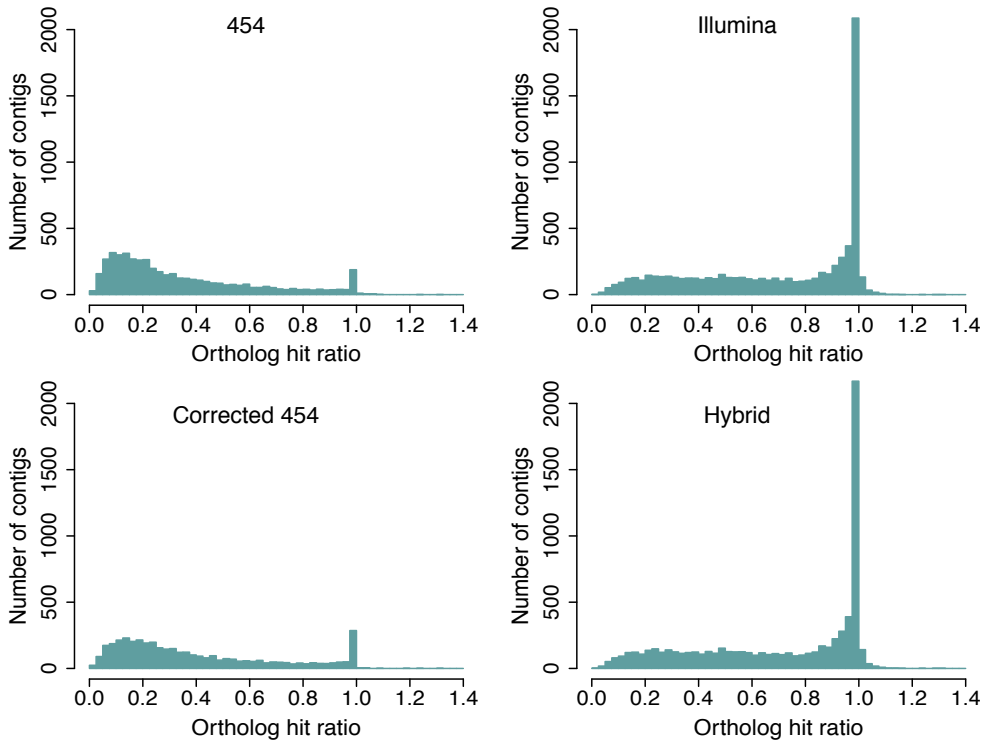

Figure 2.3: Histograms of ortholog hit ratios (*i.e.*, contig lengths relative to ortholog length) for contigs generated from the 454, Illumina, Corrected 454, and Hybrid transcriptome assemblies. Ratios equal to 1 indicate fully assembled transcripts. Values < 1 signify partial transcripts, while values greater 1 than represent contigs with insertions relative to orthologs. Orthologs were determined by 1:1 reciprocal best blast hits with chicken.

454 or Corrected 454 assemblies. The greater depth of coverage provided by Illumina sequencing may be partly responsible for the increase in the number of full-length or nearly full-length assembled transcripts. The Illumina and Hybrid assemblies performed equally well at constructing complete transcripts across both small and large orthologous genes (Figure 2.4), whereas the ability of the 454 and Corrected 454 assemblies to build full transcripts degraded quickly with ortholog length (Figure 2.4). Our results corroborate previous work demonstrating that Illumina-only assemblies produce high completeness of transcripts across many ortholog sizes (O'Neil *et al.*, 2010; Van Belleghem, *et al.,* 2012).

Independent RNA-Seq assessment

For *de novo* transcriptome assemblies to be useful for gene expression analysis, a large proportion of high quality RNA-Seq reads should map unambiguously to a single contig with few errors. We generated Illumina sequences from foam glands of an independent set of Japanese quail males and aligned reads to each of the four transcriptome assemblies and the chicken coding sequence set. To assess each assembly's utility for RNA-Seq, we calculated the total number of aligned reads and the number that mapped uniquely or ambiguously (Figure 2.5). Our results suggest that assemblies built from Illumina data alone offer the best combination of quantity (total number) and quality (proportion unique) of mapped reads for RNA-Seq. These also out-performed the chicken coding sequence set.

A large number of reads aligned to the Illumina and Hybrid assemblies, an intermediate amount to the 454 assembly, and a relatively low number to the Corrected 454 and chicken transcriptomes (Figure 2.5). Strikingly, the Illumina and 454 assemblies

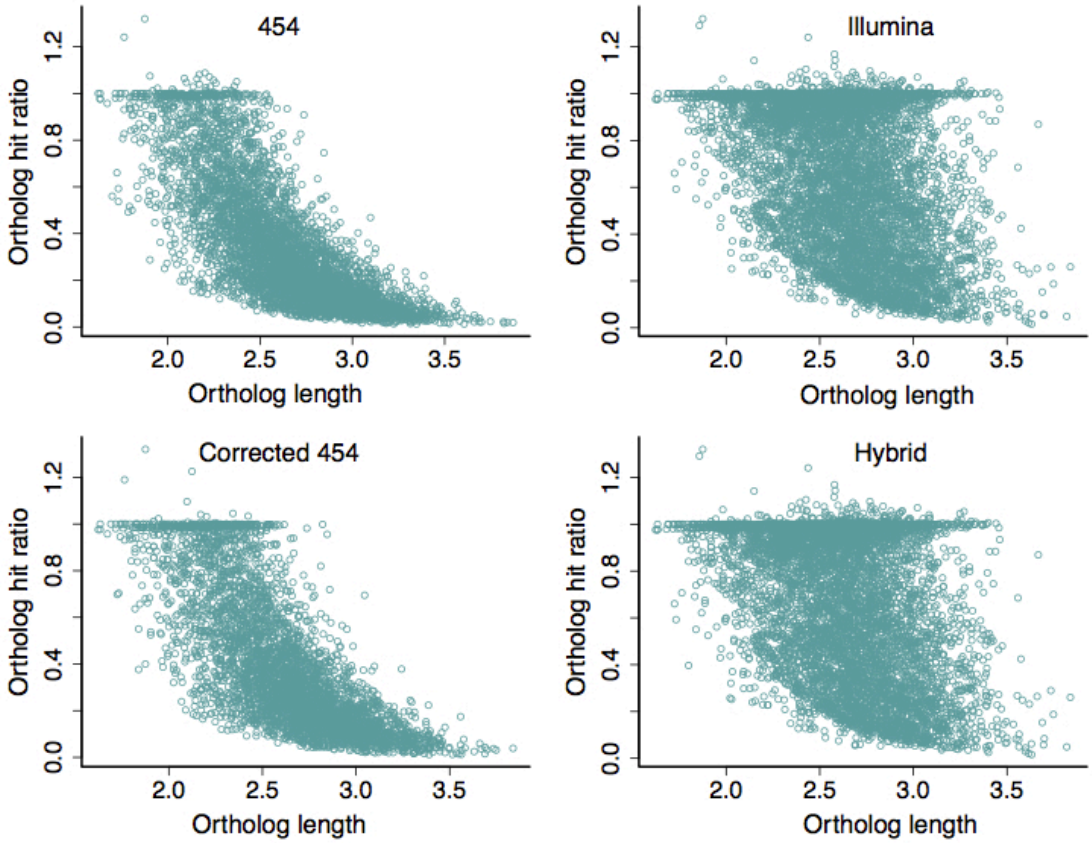

Figure 2.4: Relationship between ortholog hit ratio and ortholog length for the 454, Illumina, Corrected 454, and Hybrid assemblies. The ortholog hit ratio standardizes contig lengths relative to ortholog length. Contigs representing complete transcripts will have ratios equal to 1. Ortholog lengths are in amino acids and were $\log_{10}$ transformed. Orthologs were determined by 1:1 reciprocal best blast hits with chicken.

allowed for the lowest and highest number of uniquely mapped reads (excluding chicken; Figure 2.5). The addition of 454 data to create the Hybrid assembly increased the number of reads that mapped ambiguously, but retained the total number of reads mapped (Figure 2.5). Correcting the 454 data with Illumina short reads removed the majority of ambiguous read assignments, though at the expense of mapped reads (Figure 2.5). Reads can map to multiple contigs due to repetition in a transcriptome such as splitting of contigs that should be joined or the presence of large gene families and isoforms. The assembler used for the Illumina transcriptome, Trinity, explicitly retained isoforms, while the one used for the other transcriptomes, iAssembler, does not (Grabherr *et al.*, 2011; Zheng *et al.*, 2011). Thus, retention of paralogs of large gene families or failure to link contigs in unigenes likely explain the ambiguity present in reads mapped to the 454 and Hybrid transcriptomes. Either the nature of sequence data or the assembler could be responsible for the ambiguity.

Biases inherent to next-generation sequencing can compromise accurate quantification of gene expression (Fang & Cui, 2011). To assess potentially confounding errors in our various transcriptomes, we computed the number of insertions and deletions (*i.e.*, "indels") per 100,000 bp identified between the consensus alignments of RNA-Seq reads and each assembly (Table 2.3). One striking result was the high incidence of indels reported when examining the 454 assembly. While we anticipated a higher number of indels due to well-documented homopolymer sequencing errors intrinsic to 454 technology, the volume was surprising (Hudson, 2008; Gilles *et al.*, 2011). Mapping RNA-seq reads to the 454 assembly produced roughly two orders of magnitude higher incidence of indels than alignments with any other assembly (Table 2.3).

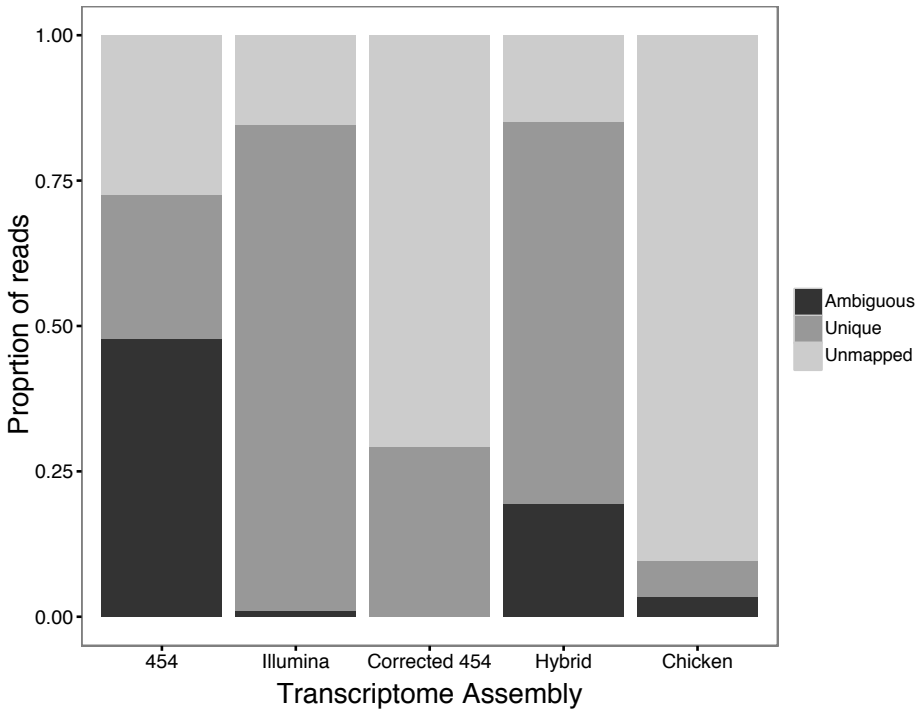

Figure 2.5: Proportion of 88,446,213 RNA-Seq reads that mapped uniquely, mapped ambiguously, or were not mapped to each of the transcriptome assemblies and chicken coding sequence.

Indels could reflect errors in the assembled transcriptome, mistakes in the RNA-Seq data, or true polymorphisms. However, we believe our 454 transcriptome assembly is at fault given the low incidence of indels from the chicken, Hybrid, and Illumina assemblies and previously discussed homopolymer issues with 454 technology (Hudson, 2008; Gilles *et al.*, 2011). False indels in transcriptomes may be problematic for RNA-Seq. They could result in a reduced number of high quality mapped reads or an increase in mis-assigned reads, both of which would negatively affect the detection of real differences in expression. Downstream applications with transcriptomes that rely on properly called open reading frames (*e.g.*, calculation of evolutionary rates) would be difficult due to erroneous frameshifts produces by false indels. Both the Hybrid and Illumina assemblies formed numbers of indels as the chicken coding sequence (Table 2.3). This is reassuring, as the chicken coding sequences are in frame (Table 2.2), suggesting that false indels are not a problem of our RNA-Seq data or our Hybrid and Illumina assemblies. It is worth noting that the quantification of differential gene expression can be robust to sequencing errors, though this finding was based on the errors and error rates specific to Illumina, not 454, sequencing (Vijay *et al.*, 2012).

Recent work suggests that directly mapping RNA-Seq reads to a related species' transcriptome (up to 15% divergent) outperforms mapping to *de novo* assembled transcriptomes in terms of accurately quantifying gene expression (Vijay *et al.*, 2012). To explore this idea, we aligned our RNA-Seq data to the chicken coding sequence set. Japanese quail and chicken experience on average 14% sequence divergence at protein-coding mitochondrial loci (Desjardins & Morais, 1991). Despite levels of divergence within the recommended range, we found that directly mapping Japanese quail RNA-Seq reads to

the chicken transcriptome performed poorly, as few reads aligned, many of which had ambiguous assignments (Figure 2.6).

Differences in the nature of the data examined may explain the disparity in the conclusions. To mimic reference-based mapping, Vijay *et al.* (2012) introduced varying levels of divergence (5-30%) *in silico* to the zebra finch transcriptome and mapped simulated RNA-Seq reads, also from zebra finch, back to the various transcriptomes. Their datasets accounted for simple differences due to nucleotide polymorphisms and indels, but did not incorporate more complex forms of variation that could affect the ability to map RNA-Seq data (*e.g.*, inversions, gene rearrangements, duplications, exon shuffling). As we utilized non-simulated data, our reference-based mapping approach encompassed both simple and complex forms of sequence divergence that occurred after the Japanese quail and chicken lineages split. Increasing transcriptome complexity (size, paralogs, alternatively spliced isoforms) negatively affects both *de* novo transcriptome assembly and the ability to quantify gene expression (Vijay *et al.*, 2012). Therefore, we caution directly mapping to a reference transcriptome from a model species, unless sequence differences between the target and reference are known to be simple and no other options are available.

Conclusions

We compared assemblies generated from mixtures of 454 and Illumina reads for *de novo* transcriptome assembly and utility for RNA-Seq analyses in a non-model species. Hybrid and Illumina assemblies produced longer contigs covering more of the transcriptome than 454-based assemblies. While the Hybrid assembly often performed the

absolute best in standard assays of transcriptome quality, improvements over the Illumina assembly were marginal. Hybrid and Illumina assemblies also afforded more gene annotations that better reproduced ortholog lengths. However, if a goal is to identify the maximum number of annotations, utilizing both 454 and Illumina is preferred, as each contributes a significant number of annotations. The Illumina assembly offered the best scaffold for RNA-Seq data, as it delivered the highest number of uniquely mapped reads. Our results may be unsurprising given the vast differences in the number of reads generated by the two technologies. However, cost is often a limiting factor when working with non-model species and we spent approximately the same amount of money on generating both types of data.

A current challenge facing the non-model community is how to navigate the landscape of next-generation sequencing efficiently and economically. In the past, researchers considered a two-step approach, first building a transcriptome (typically from 454 reads) that later served as a scaffold for mapping RNA-Seq reads, generally generated from a separate Illumina run (*e.g.*, Su *et al.*, 2011; Jensen *et al.*, 2012). From sequencing one-half of an Illumina lane, we assembled a high quality transcriptome that consistently outperformed a 454 transcriptome for less money. Further, our Illumina data averaged 37 million reads per sample, which exceeds the suggested recommended number for robust detection of differential gene expression (10-30 million reads; Wang *et al.*, 2011, but see Toung *et al.*, 2011). To be fair, our study represents a single snapshot in time and is conservative. Both sequencing platforms continue to produce more data with increasing read lengths and fewer errors, at less cost. Nevertheless, for researchers on limited budgets with few genomic resources, the present study shows that sequencing

transcriptomes with Illumina technology provides sufficient data for *de novo* assembly and RNA-Seq analysis in a single step.

Acknowledgements

We would like to thank Steve Bogdanowicz and Jennifer Mosher for support and guidance with next generation sequencing; Elizabeth Adkins-Regan for providing and housing quail; Stephanie Iacovelli, Nicole Baran, and Dave Cerasale for quail handling; Tim van Deusen, Percy Smith, Linda Vann, and Stephanie Martin for quail care; the RNA-Seq reading group for advice and discussion; and the Harrison lab plus lab 'regulars' for feedback on an earlier version of the manuscript. This research was funded by an NSF DDIG DEB-1010757 to FRF and RGH, P.E.O. Scholar, Cornell Sigma Xi, Andrew W. Mellon, and Paul. F. Feeny awards to FRF.

References

Adkins, E. 1977. Effects of diverse androgens on sexual behavior and morphology of castrated male quail. *Hormones And Behavior* **8**: 201–207.

Adkins-Regan, E. 1999. Foam produced by male *Coturnix* quail: What is its function? *Auk* **116**: 184–193.

Birzele, F., Schaub, J., Rust, W., Clemens, C., Baum, P., Kaufmann, H., *et al.* 2010. Into the unknown: expression profiling without genome sequence information in CHO by next generation sequencing. *Nucleic Acids Research* **38**: 3999–4010.

Bork, P. & Koonin, E.V. 1998. Predicting functions from protein sequences--where are the bottlenecks? *Nat Genet* **18**: 313–318.

Cahais, V., Gayral, P., Tsagkogeorga, G., Melo-Ferreira, J., Ballenghien, M., Weinert, L., *et al.* 2012. Reference-free transcriptome assembly in non-model animals from next-generation sequencing data. *Mol Ecol Resour* **12**: 834–845.

Chen, H. & Boutros, P.C. 2011. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics* **12**: 35. BioMed Central Ltd.

Chen, S., Yang, P., Jiang, F., Wei, Y., Ma, Z. & Kang, L. 2010. De novo analysis of transcriptome dynamics in the migratory locust during the development of phase traits. *PLoS ONE* **5**: e15633.

Cheng, K.M., Hickman, A.R. & Nichols, C.R. 1989a. Role of the proctodeal gland foam of male Japanese quail in natural copulations. *Auk* **106**: 279–285.

Cheng, K.M., McIntyre, R.F. & Hickman, A.R. 1989b. Proctodeal gland foam enhances competitive fertilization in domestic Japanese quail. *Auk* **106**: 286–291.

Coil, W.H. & Wetherbee, D.K. 1959. Observations on the cloacal gland of the Eurasian quail *Coturnix coturnix*. *The Ohio Journal of Science* **59**: 268–270.

Crawford, J.E., Guelbeogo, W.M., Sanou, A., Traoré, A., Vernick, K.D., Sagnon, N., *et al.* 2010. De novo transcriptome sequencing in *Anopheles funestus* using Illumina RNA-Seq technology. *PLoS ONE* **5**: e14202.

Desjardins, P. & Morais, R. 1991. Nucleotide sequence and evolution of coding and noncoding regions of a quail mitochondrial genome. *J Mol Evol* **32**: 153–161.

Ekblom, R. & Galindo, J. 2010. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* **107**: 1–15. Nature Publishing Group.

Ekblom, R., Farrell, L.L., Lank, D.B. & Burke, T. 2012. Gene expression divergence and nucleotide differentiation between males of different color morphs and mating

strategies in the ruff. *Ecology and Evolution* **2**: 2485–2505.

Ellegren, H. 2008. Sequencing goes 454 and takes large-scale genomics into the wild. *Molecular Ecology* **17**: 1629–1631.

Etebari, K., Palfreyman, R.W., Schlipalius, D., Nielsen, L.K., Glatz, R.V. & Asgari, S. 2011. Deep sequencing-based transcriptome analysis of *Plutella xylostella* larvae parasitized by *Diadegma semiclausum*. *BMC Genomics* **12**: 446. BioMed Central Ltd.

Fang, Z. & Cui, X. 2011. Design and validation issues in RNA-seq experiments. *Briefings in Bioinformatics* **12**: 280–287.

Feldmeyer, B., Wheat, C.W., Krezdorn, N., Rotter, B. & Pfenninger, M. 2011. Short read Illumina data for the de novo assembly of a non-model snail species transcriptome (*Radix balthica*, Basommatophora, Pulmonata), and a comparison of assembler performance. *BMC Genomics* **12**: 317. BioMed Central Ltd.

Finseth, F.R., Bondra, E.R., & Harrison, R.G. 2013a. Combined proteomics and RNA-Seq reveal contrasting patterns of selective constraint and rapid evolution in a novel reproductive accessory fluid. Chapter 4, Ph.D. Thesis, Cornell University.

Finseth, F.R., Bondra, E.R., & Harrison, R.G. 2013b. Phenotypic divergence arose without major protein-coding sequence divergence in a novel reproductive gland. Chapter 3, Ph.D. Thesis, Cornell University.

Finseth, F.R., Iacovelli, S.R., Harrison, R.G., & Adkins-Regan, E.K. 2013c. A non-semen copulatory fluid influences the outcome of sperm competition in Japanese quail. *Journal of Evolutionary Biology, in press.*

Fujihara, N. 1992. Accessory reproductive fluids and organs in male domestic birds. *Worlds Poultry Science Journal* **48**: 39–56.

Gilles, A., Meglécz, E., Pech, N., Ferreira, S., Malausa, T. & Martin, J.-F. 2011. Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. *BMC Genomics* **12**: 245.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., *et al.* 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* **29**: 644–652.

Hornett, E.A. & Wheat, C.W. 2012. Quantitative RNA-Seq analysis in non-model species: assessing transcriptome assemblies as a scaffold and the utility of evolutionary divergent genomic reference species. *BMC Genomics* **13**: 1–1.

Huang, Y., Niu, B., Gao, Y., Fu, L. & Li, W. 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* **26**: 680–682.

Hudson, M.E. 2008. Sequencing breakthroughs for genomic ecology and evolutionary

biology. *Mol Ecol Resour* **8**: 3–17.

Jensen, J.K., Schultink, A., Keegstra, K., Wilkerson, C.G., & Pauly, M. RNA-Seq analysis of developing Nasturtium seeds (*Tropaeolum majus*): identification and characterization of an additional galactosyltransferase involved in xyloglucan biosynthesis. *Molecular plant* **5:** 984 - 992.

Kayang, B., Fillon, V., Inoue-Murayama, M., Miwa, M., Leroux, S., Feve, K., *et al.* 2006. Integrated maps in quail (*Coturnix japonica*) confirm the high degree of synteny conservation with chicken (*Gallus gallus*) despite 35 million years of divergence. *BMC Genomics* **7**.

King, A.S. 1981. Cloaca. In: *Form and function in birds* (A. S. King & J. McLelland, eds), pp. 63–105. Academic Press, New York.

Klemm, R., Knight, C.E. & Stein, S. 1973. Gross and microscopic morphology of glandula-proctodealis (foam gland) of *Coturnix c. japonica* (aves). *J Morphol* **141**: 171–184.

Koonin, E.V. 2005. Orthologs, paralogs, and evolutionary genomics. *Annual Review of Genetics* **39**: 309–338.

Kumar, S. & Blaxter, M.L. 2010. Comparing de novo assemblers for 454 transcriptome data. *BMC Genomics* **11**: 571.

Lake, P. 1981. Male genital organs. In: *Form and function in birds* (A. S. King & J. McLelland, eds), pp. 1–61. Academic Press, London.

Li, H. & Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760.

Meyer, E., Aglyamova, G.V., Wang, S., Buchanan-Carter, J., Abrego, D., Colbourne, J.K., *et al.* 2009. Sequencing and de novo analysis of a coral larval transcriptome using 454 GSFlx. *BMC Genomics* **10**: 219.

Milano, I., Babbucci, M., Panitz, F., Ogden, R., Nielsen, R.O., Taylor, M.I., *et al.* 2011. Novel Tools for conservation genomics: comparing two high-throughput approaches for SNP discovery in the transcriptome of the European hake. *PLoS ONE* **6**: e28008.

Min, X.J., Butler, G., Storms, R. & Tsang, A. 2005. OrfPredictor: predicting protein-coding regions in EST-derived sequences. *Nucleic Acids Research* **33**: W677–W680.

O'Neil, S.T., Dzurisin, J.D., Carmichael, R.D., Lobo, N.F., Emrich, S.J. & Hellmann, J.J. 2010. Population-level transcriptome sequencing of nonmodel organisms *Erynnis propertius* and *Papilio zelicaon*. *BMC Genomics* **11**: 310.

Poiani, A. 2006. Complexity of seminal fluid: a review. *Behav Ecol Sociobiol* **60**: 289–310.

Rawat, A., Elasri, M.O., Gust, K.A., George, G., Pham, D., Scanlan, L.D., *et al.* 2012. CAPRG:

Sequence Assembling Pipeline for Next Generation Sequencing of Non-Model Organisms. *PLoS ONE* **7**: e30370.

Rawat, A., Gust, K.A., Deng, Y., Garcia-Reyero, N., Quinn, M.J., Johnson, M.S., *et al.* 2010a. From raw materials to validated system: the construction of a genomic library and microarray to interpret systemic perturbations in Northern bobwhite. *Physiological Genomics* **42**: 219–235.

Rawat, A., Gust, K.A., Elasri, M.O. & Perkins, E.J. 2010b. Quail Genomics: a knowledgebase for Northern bobwhite. *BMC Bioinformatics* **11**: S13.

Reading, B.J., Chapman, R.W., Schaff, J.E., Scholl, E.H., Opperman, C.H. & Sullivan, C.V. 2012. An ovary transcriptome for all maturational stages of the striped bass (Morone saxatilis), a highly advanced perciform fish. *BMC Research Notes* **5**: 111. BioMed Central Ltd.

Rumble, S.M., Lacroute, P., Dalca, A.V., Fiume, M., Sidow, A. & Brudno, M. 2009. SHRiMP: Accurate Mapping of Short Color-space Reads. *PLoS Comput Biol* **5**: e1000386.

Sammeth, M. 2009. Complete Alternative Splicing Events Are Bubbles in Splicing Graphs. *Journal of Computational Biology* **16**: 1117–1140.

Sasazaki, S., Hinenoya, T., Fujima, D., Kikuchi, S., Fujiwara, A. & Mannen, H. 2006a. Mapping of expressed sequence tag markers with a cDNA-amplified fragment length polymorphism method in Japanese quail (Coturnix japonica). *Animal Science Journal* **77**: 42–46.

Sasazaki, S., Hinenoya, T., Lin, B., Fujiwara, A. & Mannen, H. 2006b. A comparative map of macrochromosomes between chicken and Japanese quail based on orthologous genes. *Anim Genet* **37**: 316–320.

Schumacher, M. & Balthazart, J. 1983. The effects of testosterone and its metabolites on sexual-behavior and morphology in male and female Japanese quail. *Physiol Behav* **30**: 335–339.

Schwartz, T.S., Tae, H., Yang, Y., Mockaitis, K., Van Hemert, J.L., Proulx, S.R., *et al.* 2010. A garter snake transcriptome: pyrosequencing, de novo assembly, and sex-specific differences. *BMC Genomics* **11**: 694. BioMed Central Ltd.

Seiwert, C. & Adkins-Regan, E. 1998. The foam production system of the male Japanese quail: Characterization of structure and function. *Brain Behav Evol* **52**: 61–80.

Singh, R.P., H Sastry, von, K., Shit, N., Pandey, N.K., Singh, K.B., Mohan, J., *et al.* 2011. Cloacal gland foam enhances motility and disaggregation of spermatozoa in Japanese quail (*Coturnix japonica*). *Theriogenology* **75**: 563–569. Elsevier Inc.

Singh, R.P., Sastry, K.V.H., Pandey, N.K., Singh, K.B., Malecki, I.A., Farooq, U., *et al.* 2012. The

role of the male cloacal gland in reproductive success in Japanese quail (*Coturnix japonica*). *Reprod. Fertil. Dev.* **24**: 405.

Su, C.L., Chao, Y.T., Alex Chang, Y.C., Chen, W.C., Chen, C.Y., Lee, A.Y., *et al.* 2011. De Novo assembly of expressed transcripts and global analysis of the *Phalaenopsis aphrodite* transcriptome. *Plant and Cell Physiology* **52**: 1501–1514.

Surget-Groba, Y. & Montoya-Burgos, J.I. 2010. Optimization of de novo transcriptome assembly from next-generation sequencing data. *Genome Research* **20**: 1432–1440.

Tarazona, S., Garcia-Alcalde, F., Dopazo, J., Ferrer, A. & Conesa, A. 2011. Differential expression in RNA-seq: A matter of depth. *Genome Research* **21**: 2213–2223.

Tatusov, R.L. 1997. A genomic perspective on protein families. *Science* **278**: 631–637.

Toung, J.M., Morley, M., Li, M. & Cheung, V.G. 2011. RNA-sequence analysis of human B-cells. *Genome Research* **21**: 991–998.

Van Belleghem, S.M., Roelofs, D., Van Houdt, J. & Hendrickx, F. 2012. De novo transcriptome assembly and SNP discovery in the wing polymorphic salt marsh beetle *Pogonus chalceus* (Coleoptera, Carabidae). *PLoS ONE* **7**: e42605.

Vera, J.C., Wheat, C.W., Fescemyer, H.W., Frilander, M.J., Crawford, D.L., Hanski, I., *et al.* 2008. Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Molecular Ecology* **17**: 1636–1647.

Vijay, N., Poelstra, J.W., Künstner, A. & Wolf, J.B.W. 2012. Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive in silicoassessment of RNA-seq experiments. *Molecular Ecology* **22**:620-63434.

Wall, P.K., Leebens-Mack, J., Chanderbali, A.S., Barakat, A., Wolcott, E., Liang, H., *et al.* 2009. Comparison of next generation sequencing technologies for transcriptome characterization. *BMC Genomics* **10**: 347.

Wang, Y., Ghaffari, N., Johnson, C.D., Braga-Neto, U.M., Wang, H., Chen, R., *et al.* 2011. Evaluation of the coverage and depth of transcriptome by RNA-Seq in chickens. *BMC Bioinformatics* **12**: S5.

Wang, Z., Gerstein, M. & Snyder, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10**: 57–63.

Wheat, C.W. 2010. Rapidly developing functional genomics in ecological model systems via 454 transcriptome sequencing. *Genetica* **138**: 433–451.

Xia, Z., Xu, H., Zhai, J., Li, D., Luo, H., He, C., *et al.* 2011. RNA-Seq analysis and de novo transcriptome assembly of Hevea brasiliensis. *Plant Mol Biol* **77**: 299–308.

Xu, L. 2006. Average gene length is highly conserved in prokaryotes and eukaryotes and diverges only between the two kingdoms. *Molecular Biology and Evolution* **23**: 1107–1108.

Zheng, Y., Zhao, L., Gao, J. & Fei, Z. 2011. iAssembler: a package for de novo assembly of Roche-454/Sanger transcriptome sequences. *BMC Bioinformatics* **12**: 453. BioMed Central Ltd.

CHAPTER 3

PHENOTYPIC DIVERGENCE AROSE WITHOUT MAJOR PROTEIN-CODING SEQUENCE DIVERGENCE IN A NOVEL REPRODUCTIVE GLAND

Abstract

Sexual selection often promotes the evolution of novel adaptations that increase an organism's relative fertilization success. While it is well-documented that proteins involved with reproduction frequently diverge rapidly and sometimes adaptively, those that encode novel phenotypes may be subject to unique selective pressures. Here, we investigate the molecular evolutionary patterns of genes that encode a novel reproductive trait that mediates sperm competition, the foam gland of Japanese quail (*Coturnix japonica*). We combined an RNA-Seq and comparative genomics approach to compare pairwise and lineage-specific evolutionary rates of genes with widespread expression or those with biased expression in the foam gland, testis, or liver of Japanese quail. Overall, we identified pronounced heterogeneity in evolutionary rates. Foam gland genes evolved under strong selective constraint, while testis genes evolved rapidly and sometimes under positive selection. These striking differences were robust to variation in gene expression and promoted by the subset of genes primarily responsible for the reproductive function of the foam gland and testis. Our data also provide no evidence that foam gland genes experienced class-wide shifts of selective constraint after the divergence of quail and chicken, but we do observe a burst of accelerated evolution in testis genes along the quail lineage. Thus, the evolution of the novel foam gland proceeded without pervasive changes in protein-coding sequence, suggesting that rapid divergence is not necessarily a consequence of involvement in sexual selection.

Introduction

Across diverse species, proteins involved in sexual reproduction consistently reveal rapid sequence divergence (Swanson & Vacquier, 2002a,b; Panhuis *et al.*, 2006; Clark *et al.*,

2006; Turner & Hoekstra, 2008). This pattern is nearly universal for genes expressed in male reproductive tissues that are shared across taxa (e.g., testis, accessory glands). However, many adaptations for reproduction are novel (Eberhard, 1985; Andersson, 1994, *e.g.*, Singh & Kulathinal, 2000; Williford *et al.*, 2004; Wagner & Lynch, 2005; Schlupp *et al.*, 2010; Wiens *et al.*, 2011). Novel traits may often arise from changes in expression level, rather than protein-coding sequence (Sean Carroll *et al.*, 2001; Wray, 2007). Thus, the expected pattern of sequence divergence for proteins encoding a novel reproductive phenotype is unclear.

Genes encoding a novel reproductive trait may evolve rapidly, particularly if involved in post-mating sexual selection (PMSS). PMSS of some form (sperm competition, sexual conflict, and/or cryptic female choice) is often cited as responsible for the rapid evolution of reproductive tract genes (reviewed in Swanson & Vacquier, 2002a,b; Turner & Hoekstra, 2008; Wong, 2011). While the contribution of PMSS to the divergence of a wide range of reproductive phenotypes is clear (*e.g.,* Arnqvist, 1998; Hosken, 1998; Anderson & Dixson, 2002; Ramm & Stockley, 2010), empirical evidence for its role in the evolution of reproductive tract genes is mixed (reviewed in Wong, 2011). Investigations of single reproductive genes find some, though surprisingly few, significant associations between intensity of PMSS and the rate of molecular evolution (Dorus *et al.*, 2004; Herlyn & Zischler, 2007; Hurle *et al.*, 2007; Ramm *et al.*, 2008; Martin-Coello *et al.*, 2009; O'Connor & Mundy, 2009; Finn & Civetta, 2010). Studies adopting a multigene approach generally confirm the prediction that average evolutionary rates of reproductive genes are higher in taxa with higher levels of PMSS (Wagstaff, 2005; Kelleher *et al.*, 2007; Almeida & DeSalle, 2008; Wong, 2010, but see Walters & Harrison, 2011; Good *et al.*, 2013). Therefore, the simple

prediction is that genes encoding reproductive traits involved in PMSS will exhibit elevated rates of evolution when compared to those from non-reproductive tissues.

The simple prediction becomes less clear when the phenotype subject to PMSS is novel. Evolutionary rates are calculated by comparing sequences among species, with the assumption that genes have had similar functions in the lineages under investigation. This assumption is not necessarily valid when examining evolutionary rates for genes that contribute to novel phenotypes. Novel phenotypes often result from co-option of ancestral genes for new functional roles, which may cause a shift in the direction or degree of selection in lineage-specific ways (True & Carroll, 2002; Carroll, 2005). A co-opted gene may evolve new functions through *i)* amino acid substitutions in parts of a protein-coding sequence, *ii)* modifications to regulatory elements that result in novel spatial or temporal patterns of expression, or *iii)* divergence from an ancestral gene in either regulatory or protein-coding regions, following gene duplication (True & Carroll, 2002). Rates of protein evolution would only be elevated when protein-coding sequences change significantly after co-option. Even when relevant changes occur in protein-coding regions, if the novel function arose recently or is dependent on changes in only a few codons, there may not be sufficient signal to detect a shift in selective pressure, even if selection was strong and positive. Novel expression patterns associated with modified regulation could actually slow down protein evolution due to greater pleiotropic constraints associated with broader expression patterns (Duret & Mouchiroud, 2000; Zhang & Li, 2004; Liao, 2006; Larracuente *et al.*, 2008). It is likely that many genes are important for novel reproductive phenotypes and that these vary for acquisition time of new functions, intensity and direction of selection, and degree of selective constraint.

Here, we investigate the molecular evolution of proteins associated with the foam gland of male Japanese quail—a novel, sexually dimorphic accessory gland that is involved in reproduction and plays a role in PMSS (Finseth *et al.* 2013b). Males of the genus *Coturnix* are unique among birds in possessing a well-developed foam gland (also known as the "proctodeal gland" or "cloacal gland") (Klemm *et al.*, 1973). This large, red, external protuberance is interdigitated with the cloacal sphincter muscle and lies dorsal to the cloaca in sexually mature males (McFarland, L. Z. *et al.*, 1968; Klemm *et al.*, 1973). The foam gland is considered an 'aggregate gland' as it is comprised of many individual glandular units with apparently identical function bound by connective tissue (Klemm *et al.*, 1973). These units secrete a viscous mucoprotein that is whipped into an airy, meringue-like foam by rhythmic motions of the cloacal muscle, the frequency of which increase upon detection of a female (Klemm *et al.*, 1973; Seiwert & Adkins-Regan, 1998). During copulation, males deposit semen along with a large quantity of foam in the female reproductive tract (Klemm *et al.*, 1973).

The foam gland-foam complex is an example of a novel phenotype. Although some closely related species (*e.g.*, turkey, chicken) possess rudimentary dorsal proctodeal glands, *Coturnix* males are the only avian lineage in which the gland is reddened, swollen, and noticeably protrudes from the cloacal region (Klemm *et al.*, 1973; King, 1981; Fujihara, 1992). It is also unclear whether the rudimentary glands in other Galliformes are homologous structures (Klemm *et al.*, 1973; King, 1981; Fujihara, 1992). The foam produced by *Coturnix* males is clearly chemically unique (Fujihara, 1992), and no other avian genus, including sister genera, is known to secrete a copulatory fluid with a similar foamy quality (Klemm *et al.*, 1973). When the foam-gland complex arose is not clear, but it

likely was acquired sometime after the *Alectoris* - *Coturnix* split, which occurred approximately 30-35 million years ago (van Tuinen & Dyke, 2004).

The foam produced by the foam gland likely plays a role in PMSS. Female Japanese quail mate with multiple males and can store sperm for up to 11 days, setting the stage for sexual selection to continue after mating (Sittmann & Abplanalp, 1965; Nichols, 1991; Birkhead & Fletcher, 1994). Indeed, it has been shown that foam mediates the outcome of sperm competition by conferring fertility benefits to a focal male's sperm, rather than causing detriment to a rival's sperm (Finseth *et al.,* 2013b, Cheng *et al.*, 1989a; Adkins-Regan, 1999). It is also the case that the presence of foam *i*) improves fertilization efficiency at certain stages in a female's ovulatory cycle (Cheng *et al.*, 1989a; Adkins-Regan, 1999), *ii*) extends the window of time that a male can fertilize a female's set of eggs following a single insemination (Cheng *et al.*, 1989a; Singh *et al.*, 2012), *iii*) increases sperm motility and viability *in vitro* (Cheng *et al.*, 1989b; Singh *et al.*, 2011a), *iv*) disassembles clumped sperm (Singh *et al.*, 2011a), and *v*) improves sperm transport in the oviduct (Singh *et al.*, 2012). Additionally, males with larger foam glands are more likely to secure successful copulations, although it is not clear whether this effect is due to male-male competition, female choice, or simple differences in body size or hormonal state of males with larger glands (Singh *et al.*, 2012).

We combined an RNA-Seq and comparative genomics approach to investigate the molecular evolutionary patterns of genes with tissue-specific expression in the foam glands of male Japanese quail. We compared pairwise and branch-specific estimates of evolutionary rates among ubiquitously expressed genes and genes with biased expression in the foam gland, testis, or liver. We report pronounced heterogeneity in evolutionary rate

across reproductive tissues, with foam gland genes showing surprisingly slow rates of protein evolution, whereas testis-derived transcripts evolve rapidly and experienced a burst of accelerated evolution along the *Coturnix* branch. These patterns are robust to differences in expression level and are driven by the subset of genes responsible for the reproductive function of the two tissues. Proteins associated with sperm flagella are evolving particularly rapidly in *Coturnix* quail, which we suggest may be due to coevolution between the sperm midpiece and foam.

Methods

Subjects, sequencing, and transcriptome assembly

We made a total of two transcriptomes from two reproductive (foam gland and testis) tissues and one non-reproductive (liver) tissue from six Japanese quail males. The transcriptomes included an 'exhaustive' one that was large but may have retained assembly artifacts and a 'filtered' one that consisted of the subset we had confidence represented real genes. Most downstream analyses deal only with the 'filtered' transcriptome. Details regarding subjects, library preparation, Illumina sequencing, transcriptome assembly and transcriptome characterization are in the supplementary methods (Methods S3.1).

Tissue-specific gene expression

We used an RNA-Seq approach with 18 samples (6 males x 3 tissues) to assign tissue-specificity for a given gene. We designated tissue specificity of genes two ways, '*tissue-enriched*' and '*tissue-restricted*'.

Tissue-enriched genes

A gene was considered 'enriched' when its expression levels were significantly higher in one tissue compared with the other two. We tested for differential expression of genes using the multifactor generalized linear models (glms) approach in EdgeR version 3.0.8 (Robinson *et al.*, 2010). Briefly, we normalized our RNA-Seq data using the TMM approach (Robinson & Oshlack, 2010). We fit negative binomial glms with Cox-Reid tagwise dispersion estimates to models that included tissue and male ID as factors, as our experiment was paired by subject (3 tissue samples per male). Our design matrix specified contrasts to analyze expression level in one tissue versus the average of the other two. To determine differential expression, we performed likelihood ratio tests by dropping one coefficient from the design matrix (*i.e.*, the 'null' model) and comparing it to the full model (McCarthy *et al.*, 2012). The false discovery rate was used to account for multiple testing and a cutoff of 0.05 was applied to consider a gene 'significantly' differentially expressed in a particular tissue (Benjamini & Hochberg, 1995). The list of genes significantly upregulated by more than log 2 fold change in each tissue were designated as '*tissue-enriched*'. The genes qualifying as enriched in two tissues were not included in further analyses (N = 366 out of 11,366 enriched genes). Genes not enriched for any of the three tissues comprised a fourth category termed '*ubiquitous*'.

Tissue-restricted genes

For each gene, we computed a normalized number of reads statistic (NNR), found by standardizing the number of uniquely mapped reads per gene by gene length and scaled library size (library size per sample normalized by maximum library size). Library size

was calculated in EdgeR version 3.0.8 (Robinson *et al.*, 2010). We computed the tissue-specific means of NNR for genes from both the exhaustive and filtered transcriptomes and plotted their distributions (see supplementary *Methods S3.1* for transcriptome descriptions). The first quartile NNR value generated from the exhaustive transcriptome (0.0003) was used as a cutoff value for considering a gene 'expressed' based on the NNRs estimated from the filtered transcriptome. In other words, a gene was considered expressed in a particular tissue if its average NNR for that tissue (computed from the filtered dataset) was larger than or equal to 0.0003. A gene exhibited '*tissue-restricted*' expression when it was expressed in only one tissue, whereas genes expressed in all three tissues were classed as '*ubiquitous*'. Genes considered expressed in two tissues were not included in further analyses (N = 4,534 out of 23,656 expressed genes). Therefore, the '*ubiquitous*' categories are not exactly the same for the *'tissue-restricted'* and '*tissue-enriched*' classifications. Analyses that use the above criteria are referred to as '*tissue-restricted*' throughout.

Up-regulation of genes in reproductively active foam glands and testes

The foam gland and testis are androgen dependent, and light treatments that mimic winter conditions ('short days') cause the two glands to regress and cease production of foam or sperm (Sachs, 1967). We exposed six males to short day treatments (8L: 16D light: dark cycle) in parallel with the experiments described under Methods: *Subjects, sequencing, and transcriptome assembly* and in the supplementary methods (Methods S3.1). We confirmed that short day males had regressed foam glands and testes and did not produce foam or sperm (details in Finseth *et al*., 2013a). RNA-Seq was conducted on the foam

glands and testes from these short day males, therefore sequencing an additional 12 cDNA libraries (6 short day males x 2 tissues). Genes significantly up-regulated in the foam glands and testes of long day males (*i.e.*, those from Methods: *Subjects, sequencing, and transcriptome assembly*) relative to those tissues in short day males were identified in EdgeR using the same methodology described in Methods: '*Tissue - enriched*' genes. Library preparation, Illumina sequencing, read processing and transcriptome mapping were conducted following the same procedures described in Methods S3.1. All experiments and sequencing were performed in parallel with those from Methods: *Tissue-specific gene expression* for a total of 30 cDNA libraries randomly distributed across three Illumina lanes.

Orthology assignment

To identify orthologs of quail genes in chicken (*Gallus gallus*), we obtained all Ensembl chicken proteins and coding sequences (Ensembl version 69: *G. gallus* assembly WASHUC2) via BioMart (www.biomart.org). We filtered the *G. gallus* protein set to remove redundant entries (*i.e.*, duplicates, alternative splice variants) by self-BLAST following Hornett and Wheat (2012). We identified the longest open reading frames and associated protein translations of our exhaustive transcriptome with OrfPredictor (Min *et al.*, 2005). We then identified quail:chicken orthologs via the reciprocal best blast method (Tatusov, 1997; Bork & Koonin, 1998; Koonin, 2005). In short, using BLASTp, we compared the proteins predicted from the translation of our transcriptome with the chicken protein sequences, with a cutoff e-value of $1 * 10^{-6}$ and vice versa. Orthologs were called when the top chicken hit (based on bit score) from the quail to chicken BLAST returned the original quail query in the chicken to quail BLAST (10,129 orthologs from the exhaustive

transcriptome, 9,620 from the filtered transcriptome). We restricted all downstream analyses to the subset of genes (8,668 orthologs) without multiple stop codons in the chicken coding sequences as identified by custom Perl scripts.

Pairwise estimation of evolutionary rates

Protein sequences from the 1:1 quail:chicken orthologs were aligned with CLUSTAL W version 2.1 (Larkin *et al.*, 2007), which guided alignments of their corresponding DNA sequences with PAL2NAL as implemented in the ParaAT (Parallel Alignment and Translation) tool (Zhang *et al.*, 2012). From these alignments, we estimated the ratio of the nonsynonymous substitution rate (d_N) to the synonymous substitution rate (d_S). The d_N/d_S ratio, or ω, is the normalized amino acid substitution rate and we interpret this as the 'evolutionary rate' of proteins. We calculated pairwise ω values using the model averaging approach in the software package KaKs_Calculator version 1.2 (Zhang *et al.*, 2006). Model averaging in KaKs_Calculator implements a set of 14 candidate models in a maximum likelihood framework and computes the weighted average of d_N, d_S, and ω for each gene. We removed 459 orthologs where $d_S \geq 0.48998$ (two times the mean d_S estimated from all orthologs), as this could indicate poor alignment, and three orthologs where d_S estimates approached zero resulting in an artificially inflated ω (~50). As further quality control, we confirmed that all alignments were at least 100 amino acids long and that no d_N values were greater than 1 (indicating more than 1 nonsynonymous site per nonsynonymous substitution). For d_N, d_S, and ω of the remaining 8,238 orthologs, we constructed 95% CIs by performing 10,000 bootstrap resamplings of tissue-specific means without assuming

normality. We compared the proportion of genes under positive selection ($\omega > 1$) between tissues classes within each designation (i.e., tissue-restricted or tissue-enriched).

An additional objective was to examine whether genes involved in the reproductive function of foam gland and testis are responsible for tissue-specific patterns of protein evolution. To this end, we compared average ω values between foam gland- and testis-enriched genes that were either up-regulated when the glands were in their breeding (*i.e.,* reproductive) condition or not. Manipulation of reproductive condition and determination of up-regulated genes is described in Methods: *Up-regulation of genes in reproductively active foam glands and testes*. We excluded those genes that were significantly up-regulated in both the testis and foam gland.

Partial correlation of evolutionary rate, expression level, and enrichment

Expression level is a strong negative predictor of the rate of protein evolution (Pál *et al.*, 2001; Drummond *et al.*, 2005; Lemos *et al.*, 2005; Larracuente *et al.*, 2008). To evaluate whether the trends in evolutionary rates were robust to differences in expression, we performed partial correlations between evolutionary rate, expression level, and tissue enrichment. Spearman's ρ was used to estimate pairwise correlations between quail:chicken pairwise evolutionary rates (ω), enrichment, and expression level for each tissue class. Enrichment in each tissue was determined by log fold change of a gene compared to the average of the other two calculated as described under Methods: *'Tissue - enriched*' genes. For each gene, expression levels were estimated as reads per kilobase per million mapped reads (RPKM; Mortazavi *et al.*, 2008). Partial correlations were computed from the pairwise correlation coefficients in the R packages corpcor version 1.6.5 (Schäfer

& Strimmer, n.d.). The 95% CIs of each partial correlation were approximated by bootstrapping the data 1,000 times with the R package boot version 1.3.7.

Lineage-specific shifts in evolutionary rates

To test for variation in selective pressure along a specific lineage, we compared 'branch' models in a maximum likelihood framework implemented in the CODEML program of the software package PAML version 4.7 (Z. Yang, 2007). First, we identified 1:1:1:1 orthologs between quail: chicken: zebra finch: and turkey. We downloaded Ensembl turkey (*Meleagris gallopavo* assembly UMD2) and zebra finch (*Taeniopygia guttata* assembly taeGut3.2.4) protein and coding sequences from BioMart (www.biomart.org; Ensembl version 69). Ortholog determination and sequence alignment were performed as described above (N = 5,281 orthologous groups). CODEML was used to compare the likelihood of null models where the intensity of selection (ω) was the same along all branches in the tree (*i.e.*, one-ratio models) to two alternative models: one where the quail terminal lineage has one ω (ω_{quail}) and the rest of the tree has another (ω_{tree}), and one where the chicken terminal lineage has one ω ($\omega_{chicken}$) and the rest of the tree has another. We estimated ω along the phylogeny (((quail, chicken), turkey), zebra finch) with Galliformes relationships deduced from the well-supported, multilocus phylogeny of Kimball and Braun (2008). Significance was determined by likelihood ratio tests fitted to χ^2 distributions. We corrected for multiple testing by applying the false discovery rate with a cutoff value of 0.05 for calling significance (Wang *et al.*, 2009; Benjamini & Hochberg, 1995). In 88 cases where the same gene was identified as having a 'significant' shift in

selective pressure along both the quail and chicken branches, we chose the most likely scenario according to log likelihood scores.

We predicted we would see proportionally more significant lineage-specific shifts (accelerations: $\omega_{lineage} > \omega_{tree}$ and/or decelerations: $\omega_{lineage} < \omega_{tree}$) in ω estimates from quail versus chicken foam gland genes, but approximately equal values for testis and liver genes. Therefore, we compared the proportion of genes with significant shifts in ω for a particular tissue's gene set to the transcriptome-wide expectation from either quail or chicken. The subset of genes from the tissue in question was subtracted from the transcriptome wide-numbers prior to comparison. Significance was determined with Pearson's χ^2 test without Yates' continuity correction. The false discovery rate was applied at a cutoff of 0.05 to correct for multiple tests (Benjamini & Hochberg, 1995). To examine whether the magnitude of shifts varied across tissues, we regressed lineage-specific ω estimates on ω values generated for the rest of the tree and tested whether the residuals from this model differed significantly according to tissue designation or species. We constructed 95% CIs by 10,000 bootstrap resamplings of tissue-specific means without assuming normality.

To explore gene function of genes possibly under positive selection in the quail, we annotated those genes with significant accelerations in ω_{quail} versus ω_{tree} and with $\omega_{quail} > 1$. For these genes, we determined spatial expression of the *G. gallus* ortholog in the chicken based on either Unigene expressed sequence tag (EST) data (http://www.ncbi.nlm.nih.gov/unigene) or RNA-Seq data from Ensembl (http://useast.ensembl.org/index.html). While values of $\omega > 1$ indicate positive selection, estimations of ω over entire genes are quite conservative. More powerful inferences of positive selection can be obtained by performing multiple tests that explicitly examine

whether particular codon positions are associated with $\omega > 1$ along the branch leading to quail (Z. Yang *et al.*, 2000; Anisimova *et al.*, 2002). We refrained from doing such tests here because a low number of species reduces the power and accuracy of these analyses, and well-annotated large sequence datasets currently exist for only three avian species other than quail (Anisimova *et al.*, 2001; 2002).

Data analysis

All analyses were performed in R version 2.15.2 (R Core Team 2012) unless otherwise stated. CIs were generated in the Hmisc package of R (Harrell, 2012) unless otherwise stated. All BLAST steps were performed in parallel via Cornell University's Computational Biology Application Suite for High Performance Computing (biohpc.org). Venn diagrams were made in the VennDiagram version 1.5.1 package in R (Chen & Boutros, 2011).

Results

Summary of transcriptome

Table S3.1 reports the average and standard deviation of the number of raw, filtered, mapped, and uniquely mapped reads. Table S3.2 reports standard transcriptome quality metrics including average transcript length, median transcript length, N50, and sum of transcript lengths. The majority of the filtered transcriptome was represented by full-length or nearly full-length transcripts as identified by the ortholog hit ratio (ratios = 1 in Figure S3.1). Figure S3.2 depicts the distribution of the NNR statistics for each tissue class,

which was used to assign tissue-restricted status. Samples from the three different tissues displayed distinct expression profiles (Figure S3.3). Tissue was a proportionally stronger effect than subject (*i.e.,* individual, as different tissue samples from the same males did not tend to cluster together (Figure S3.3).

The distribution of genes with tissue-enriched and tissue-restricted designations is shown in Figure 3.1. In all cases, > 99.5% of genes classed as tissue-restricted in a given tissue were also enriched in that same tissue. Tissues varied in the number of genes assigned to them, with testis having the most and the foam gland the least number of enriched and restricted genes (Figure 3.1). These differences were not due to differences in library size as both designations account for library sizes. For every gene tested, the qPCR data validated that the genes were significantly differentially up-regulated in the expected tissue (Table S3.3).

Pairwise estimation of evolutionary rates

Genes with biased expression in the foam gland evolved more slowly, while testis genes evolved more rapidly, than genes from other tissues (non-overlapping CIs, Table 3.1). Genes enriched in the foam gland averaged slower rates of protein evolution (ω) than those from testis or liver. Those genes restricted to the foam gland also evolved more slowly than similar genes in the testis. Testis-enriched genes had higher ω estimates than genes enriched in other tissues or ubiquitously expressed genes, and testis-restricted genes had greater or nearly greater evolutionary rates than those restricted to other single tissues. Genes more narrowly expressed in the testis (*i.e.*, testis-restricted) evolved faster than those only enriched in the testis (Table 3.1). Evolutionary rate did not differ with

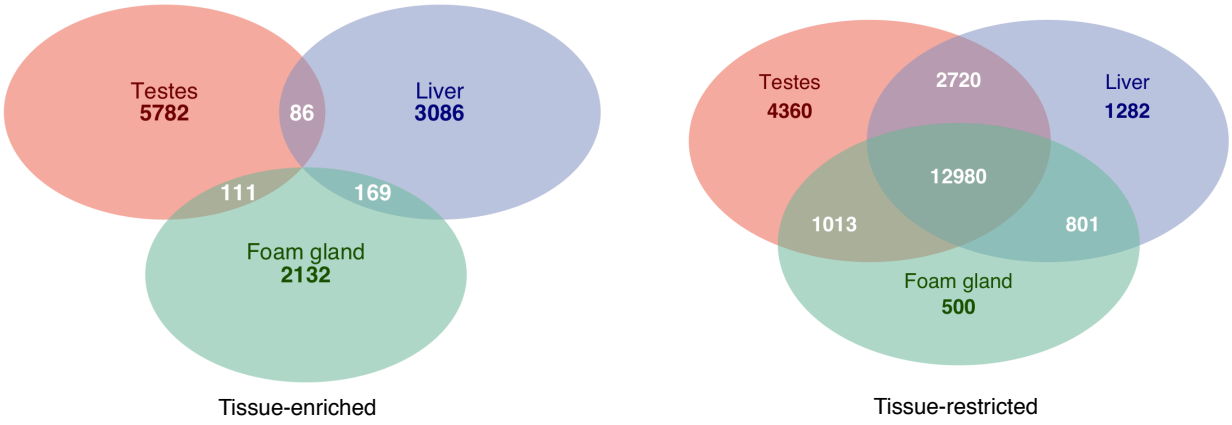

Figure 3.1. Venn diagrams depicting the numbers of genes from various tissue classes. *Tissue-enriched genes*: The number of genes enriched by at least log 2 fold in one tissue versus the other two. There are 2100 expressed genes that were not enriched in any tissue and were considered 'ubiquitous' in the tissue-enriched classification (not shown). *Tissue-restricted genes*: The number of genes expressed in a particular tissue. The 12,980 genes expressed in all three tissues are considered 'ubiquitous' in the tissue-restricted classification.

expression specificity among quail or liver genes (Table 3.1). Significantly more testis-restricted genes were under positive selection ($\omega > 1$) than were ubiquitous genes (two-tailed Fisher's exact test, $P = 7.2 * 10^{-5}$). No other tissue classes differed from each other in the number of genes under positive selection (two-tailed Fisher's exact test, $P > 0.05$ in all cases).

Patterns of variation in ω could be produced by differences in either d_N or d_S. While we observed significant differences in d_N among tissue classes consistent with trends in ω, d_S was significantly higher in genes enriched in the liver versus the other two tissues (Table 3.1). To confirm that the observed ω patterns were driven by differences in d_N, we regressed d_N on d_S and evaluated whether the residuals varied according to tissue class. This analysis suggests that the ω findings are robust to differences in d_S, as average residual differences across tissue classes replicate major patterns from pairwise ω estimates (Figure S3.4).

Genes that play a role in the reproductive function of foam gland and testis drive the major differences in evolutionary rates between the two tissues (Table 3.1). Genes up-regulated in the testis when in the breeding condition (*i.e.*, are 'active') show elevated rates of protein evolution when compared to those up-regulated in active foam glands. Conversely, testis- and foam gland-enriched genes that do not increase in expression level when in the breeding condition have no difference in evolutionary rates.

Tissue-specific trends in ω after correcting for differences in expression

Expression level is a strong negative correlate of the evolutionary rate of genes (Pál *et al.*, 2001; Drummond *et al.*, 2005; Lemos *et al.*, 2005; Larracuente *et al.*, 2008), yet most

Table 3.1 Patterns of protein evolution derived from 1:1 quail:chicken orthologs

Designation	Tissue	*N*	d_N (95% CI)[a]	d_S (95% CI)	ω (95% CI)	ω > 1[b]
Tissue-enriched						
	Foam gland	705	0.024 (0.022 - 0.026)	0.153 (0.147 - 0.160)	0.154 (0.142 - 0.167)	2
	Testis	1858	0.035 (0.033 - 0.037)	0.150 (0.147 - 0.153)	0.226 (0.216 - 0.236)	16
	Liver	848	0.030 (0.028 - 0.033)	0.166 (0.160 - 0.172)	0.187 (0.144 - 0.200)	5
	Ubiquitous	506	0.024 (0.022 - 0.027)	0.156 (0.149 - 0.162)	0.150 (0.137 - 0.167)	1
Tissue-restricted						
	Foam gland	91	0.032 (0.025 - 0.039)	0.171 (0.151 - 0.192)	0.180 (0.145 - 0.220)	0
	Testis	772	0.042 (0.039 - 0.049)	0.152 (0.147 - 0.157)	0.271 (0.253 - 0.290)	**11**[c]
	Liver	166	0.034 (0.030 - 0.040)	0.168 (0.157 - 0.180)	0.187 (0.192 - 0.256)	1
	Ubiquitous	5972	0.023 (0.022 - 0.023)	0.154 (0.152 - 0.156)	0.149 (0.136 - 0.144)	10
When sexually active[d]						
	Foam gland, up	326	0.021 (0.019 - 0.024)	0.151 (0.144 - 0.160)	0.144 (0.128 - 0.163)	1
	Testis, up	1462	0.038 (0.036 - 0.040)	0.152 (0.149 - 0.156)	0.241 (0.229 - 0.252)	14
	Foam gland, other	324	0.025 (0.022 - 0.029)	0.156 (0.147 - 0.166)	0.161 (0.142 - 0.180)	1
	Testis, other	239	0.027 (0.023 - 0.311)	0.144 (0.136 - 0.153)	0.167 (0.148 - 0.190)	1

[a]All 95% confidence generated from 10,000 bootstrap resamplings of the mean without assuming normality

[b]Number of genes with ω > 1

[c]Significantly different from tissue-restricted 'ubiquitous' genes; two-tailed Fisher's exact test P <0.001 after correcting for multiple tests

[d]The subset of 'tissue-enriched' transcripts from testis or the foam gland that are either up-regulated ('up') or not ('other') when the tissues are producing foam or sperm; Genes upregulated in both the testis and foam are not included

studies examining molecular evolutionary patterns of reproductive proteins do not correct for this (but see Meisel, 2011). We recovered the expected negative correlation between expression level and evolutionary rate in all three tissues, and find that the observed variation in rates of protein evolution (ω or d_N/d_S) across tissues is robust to differences in expression level (Figure 3.2). Enrichment in both the foam gland and liver were negatively correlated ($\rho < 0$) with the rate of protein evolution (ω or d_N/d_S), but to a significantly higher degree in the foam gland than liver (Figure 3.2). Enrichment in testis was positively correlated ($\rho > 0$) with evolutionary rate after correcting for expression levels. In all tissue classes, the rate of protein evolution and expression level were negatively correlated ($\rho < 0$) after correcting for the effects of enrichment.

Lineage-specific shifts in evolutionary rates

Genes enriched in the testis experienced proportionally more accelerations along the quail lineage ($\omega_{quail} > \omega_{tree}$) than expected based on transcriptome-wide values from quail (Pearson's χ^2, $P = 6.01 * 10^{-4}$; Figure 3.3). After correcting for multiple comparisons, the proportion of tissue-specific genes experiencing a change in selective pressure along either the quail or chicken lineage did not differ significantly from transcriptome expectations for any other subset of genes ($P > 0.05$ in all cases). The magnitude of the shifts in selective pressure also varied across tissues. Genes that were either enriched or restricted in the testis displayed larger than expected shift in selective pressure when regressing ω_{quail} on ω_{tree} (CIs > 0; Figure 3.4). These were also significantly larger than any tissue examined along the quail lineage, and greater (tissue-enriched) or nearly greater (tissue-restricted) than testis genes from chicken (Figure 3.4). These values suggest a

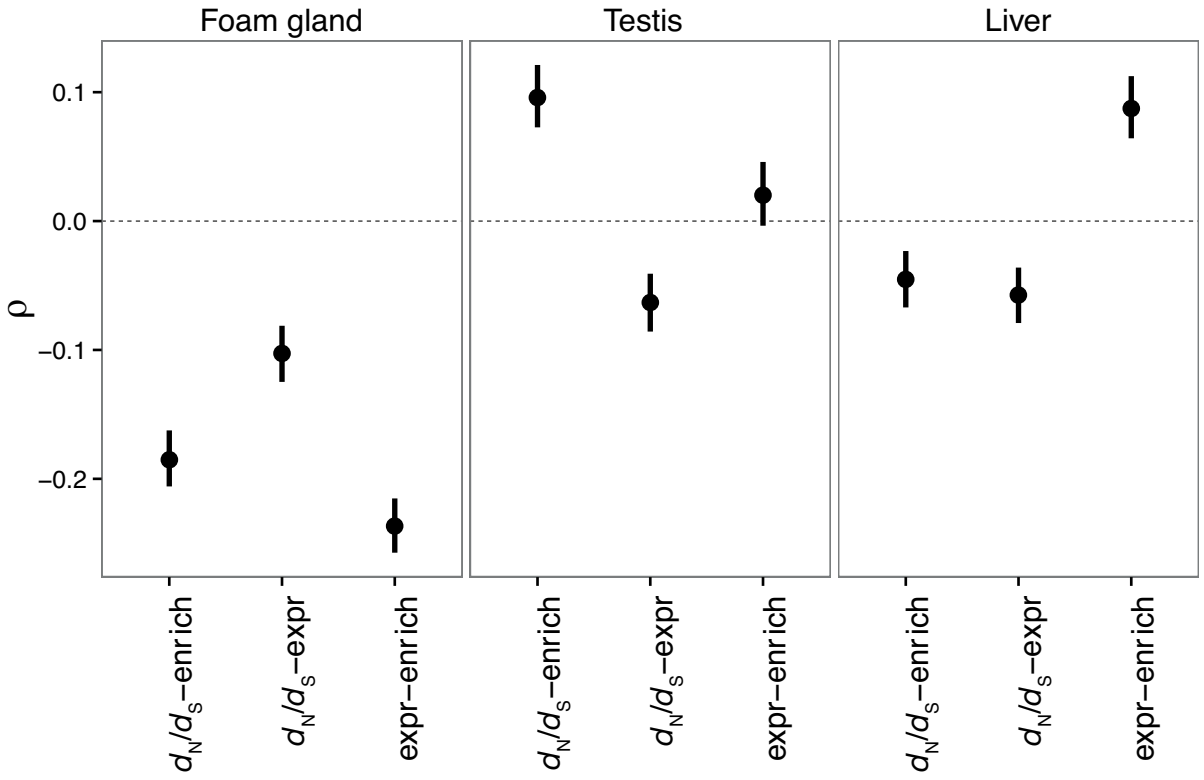

Figure 3.2. Point estimates and 95% CIs (bars) of the partial correlation coefficient (ρ) between chicken:quail evolutionary rate (d_N/d_S), gene expression level (expr) estimated by RPKM, and tissue-enrichment (enrich) measured by log fold change of a gene in a particular tissue when compared to the average of the other two. CIs were constructed from 1,000 bootstrapping resamplings of the mean without assuming normality. Sample size is 8,238 for all classes.

trend whereby genes with limited expression in the testis had larger shifts than those only enriched in the testis.

Five quail genes with significant accelerations in evolutionary rate along the quail lineage also displayed $\omega_{quail} > 1$ (Table 3.2). Four of these five genes exhibited testis-specific expression, one had foam gland-specific expression, and none were from the liver. All testis-enriched genes were also expressed in chicken testis in at least one of the chicken expression datasets.

Discussion

Morphological and functional divergence without sequence divergence

Since divergence of the quail and chicken lineages, change along the *Coturnix* branch has resulted in the appearance of an altogether new male reproductive organ. The foam gland of Japanese quail represents major phenotypic divergence along at least two axes: morphology (the gland is an enlarged, reddened protuberance consisting of many glandular units interdigitated with the cloacal sphincter muscle, the likes of which are not found in other genera (McFarland *et al.*, 1968; Klemm *et al.*, 1973)) and physiology (the gland secretes a chemically unique mucoprotein (Fujihara, 1992)). In this study, we characterized genes expressed in the foam gland and asked to what extent divergence in protein-coding sequence accompanied divergence in phenotype.

Table 3.2 Evolutionary patterns and descriptions of accelerated genes along the quail lineage where $\omega_{quail} > 1$.

Quail transcript	Chicken ortholog name	d_N[a]	d_S[a]	ω[a]	Length[a]	Divergence[a]	ω_{tree}[b]	ω_{quail}[b]	P[c]	Restricted[d]	Enriched[d]	Log FC[e]	Chicken expression[f]	Chicken expression[g]	Ensembl gene ID[h]
Coja22788_c0_seq1	Radial spoke head 1 homolog (*RSPH1*)	0.151	0.204	0.743	774	0.163	0.247	1.249	0.026	Testis	Testis	8.836	Testis, restricted	Testis, macrophage	ENSGALG00000016175
Coja15527_c0_seq1	Cancer susceptibility candidate 1 (*CASC1*)	0.120	0.133	0.905	2139	0.123	0.443	1.167	0.002	None	Testis	4.336	Wide	Testis, macrophage	ENSGALG00000014030
Coja19413_c0_seq1	Dynein, axonemal, assembly factor 1 *(DNAAF1*)	0.021	0.138	0.153	753	0.049	0.184	1.017	0.001	None	Testis	7.712	Testis, wide	Testis, wide	ENSGALG00000003258
Coja35683_c0_seq1	Pyridoxal-dependent decarboxylase domain containing 1 (*PXDC 1*)	0.041	0.093	0.441	1482	0.054	0.177	1.011	0.000	Ubiquitous	Foam gland	2.355	Wide	Wide	ENSGALG00000002796
Coja15477_c0_seq1	K(lysine) acetyltransferase 6A (*KAT6*)	0.013	0.122	0.110	1974	0.039	0.115	1.008	0.000	Ubiquitous	Testis	2.474	Spleen	Testis, wide	ENSGALG00000003641

[a]Pairwise estimates from 1:1 chicken orthologs

[b]Estimated from quail lineage specific analyses.

[c]*P* value from LRT of one versus two-ratio branch models after applying a FDR cutoff of 0.05

[d]Tissue expression in quail designated in this study

[e]Log fold change of transcript in enriched tissue versus average of other two

[f]Derived from Unigene EST data

[g]Derived from Ensembl RNA-Seq data

[h]Based on chicken ortholog

Our results provide evidence that phenotypic change in the foam gland arose without pervasive change in protein-coding genes. We compared pairwise estimates of the rate of protein evolution (ω) across genes with biased expression in the foam gland, testis, and liver. We see extreme selective constraint in foam gland genes, even after correcting for differences in expression level (Table 3.1, Figure 3.2). Surprisingly, the degree of constraint was similar to that of ubiquitously expressed genes, which are generally among the slowest evolving genes in the genome (Duret & Mouchiroud, 2000; Zhang & Li, 2004; Larracuente *et al.*, 2008). In fact, the more a gene is enriched in the foam gland, the more slowly it evolves (Figure 3.2), and the genes involved in foam production are responsible for the reduced rates of protein evolution (Table 3.1).

Pairwise estimates average ω over the entire evolutionary history separating two species. However, the intensity of selection can change over time, resulting in variation in ω along specific branches of a phylogenetic tree. This is particularly relevant with genes encoding novel structures as they may adapt to new functional roles or change in constraint. We hypothesized that, as a class, genes expressed in the foam gland experienced an increase in changes in selective pressure along the *Coturnix* branch with the acquisition of novel functions associated with foam gland formation and foam production. We also expected that we would observe relatively larger shifts in ω of foam gland genes along the quail lineage than those genes from other categories. Contrary to our predictions, the foam gland evolved without major, class-wide shifts in selective pressures along the quail lineage. We find no evidence of either excessive accelerations (*i.e.*, relaxation of selective constraint or adaptive evolution) or decelerations (*i.e.*, increased purifying selection, as may occur with additional pleiotropic constraints) of foam gland genes in

either chicken or quail (Figure 3.3). Additionally, the magnitude of shifts in selective pressure on foam gland genes did not vary from transcriptome-wide expectations (Figure 3.4).

Why are genes expressed primarily in the foam gland evolving so slowly? Many hypotheses for the rapid evolution of reproductive genes invoke antagonistic coevolutionary dynamics. For example, antagonistic interactions between the sperm of two competing males, between males and females over polyspermy conflict, or between host and pathogens have been suggested to drive reproductive protein evolution (Swanson & Vacquier, 2002a,b). While foam mediates the outcome of sperm competition, it does so through conferring benefits to a male's own sperm, rather than negatively affecting a rival's (Finseth *et al*., 2013b, Adkins-Regan, 1999). Foam also enhances sperm motility whether foam is self-derived or from a different male (Cheng *et al.*, 1989b). Thus, foam may not be under the antagonistic selective pressures generally promoted by sperm competition that fuel rapid evolution. The observed slow rates of protein evolution may be a byproduct of a lack of intrasexual conflict acting on foam gland genes. However, foam could be involved in conflict beyond sperm competition through contributing costs to females or in host-pathogen dynamics.

Foam gland proteins could evolve slowly because divergence in expression level, not protein-coding sequence, may be primarily responsible for the emergence of the foam gland. Novel phenotypes often arise from mutations affecting *cis*-regulatory elements (Carroll, 2005; Wray, 2007; Carroll, 2008). In general, PMSS may target regulation of reproductive proteins as expression of male-derived reproductive proteins evolves rapidly (Nuzhdin, 2004; Khaitovich, 2005; Ellegren & Parsch, 2007) and rate of production of

sperm proteins is a consequence of sperm competition (Ramm & Stockley, 2010). Adaptive changes in traits driven by PMSS could therefore accumulate in regulatory, rather than protein-coding, regions. Consistent with this, Martin-Coello *et al.,* (2009) showed that sexual selection favored more rapid divergence in promoter than protein-coding sequence of a sperm gene. Our study focused on protein-coding sequences and could have missed informative patterns at regulatory sequences. Explorations into expression level and regulatory divergence of foam gland genes could offer insight into how this gland evolved. The fact that the foam gland is not sex-limited could provide an additional explanation for slow rates of foam gland protein evolution (discussed in Discussion: *Testis-expressed proteins evolve rapidly*).

Some genes important in the evolution of the foam gland may not have been captured by our analyses. These would include genes that are so rapidly evolving that we could not identify orthologs, genes that underwent gene duplication since splitting with the chicken (our 1:1 reciprocal best blast approach does not detect paralogs), alternatively spliced isoforms, or new genes (Long *et al.*, 2003). Importantly, the observed negative correlation between enrichment in the foam gland and evolutionary rate is independent of any confounding ortholog-detection issues (Figure 3.2). Therefore, our finding of slow evolution of foam gland expressed genes likely reflects a real biological pattern. Of course, it is always the case that analyses of entire classes of genes or proteins may obscure interesting patterns at one or a few genes.

Heterogeneity of evolutionary rate across male reproductive tissues

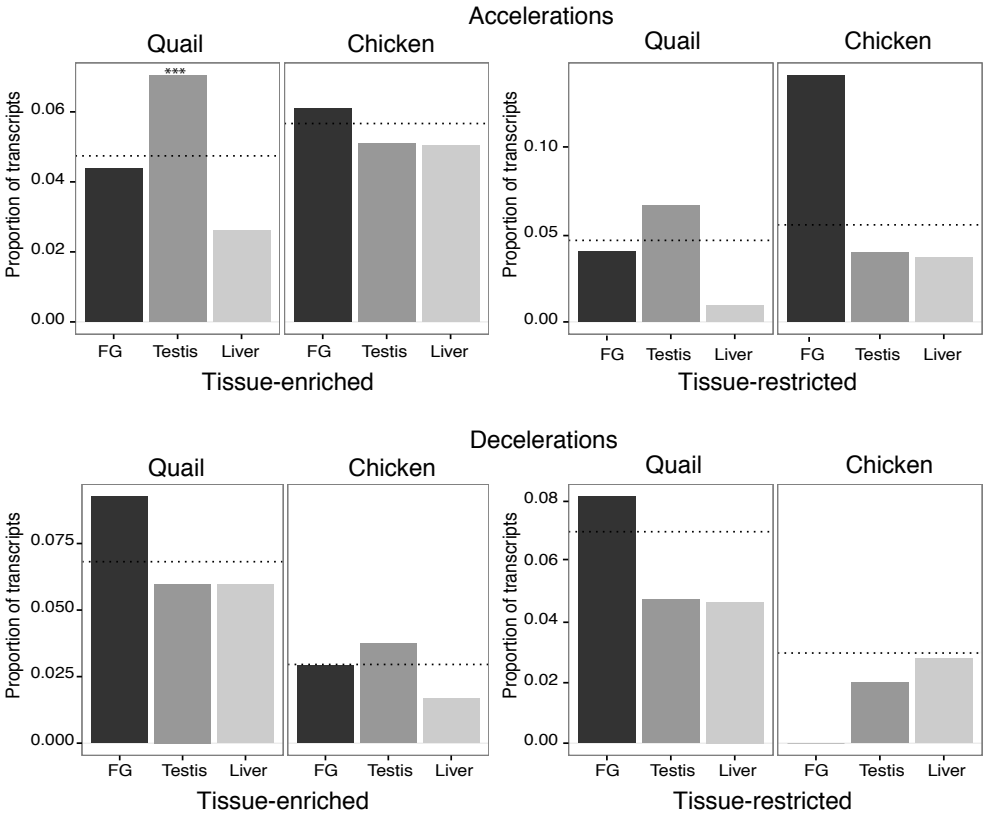

Figure 3.3. Variation in selective constraint along either the quail or chicken lineage among genes with different tissue-specific expression patterns. The proportion of genes experiencing accelerations (top two panels) or decelerations (bottom two panels) in ω estimates along either the quail (left in each panel) or chicken (right in each panel) branches when compared to the rest of the tree. Estimates are based on 1:1:1:1 orthologs between quail, chicken, turkey, and zebra finch. Proportions from genes with tissue-enrichment in the foam gland (FG), testis, or liver are on the left side, while those exhibiting tissue-restricted expression are on the right side. Tissue assignments are based on expression patterns in quail. Each subset was compared to the species-specific transcriptome wide expectation for a particular designation (dotted lines) minus the subset under investigation. Asterisks indicate a significant difference in a χ^2 test after correcting for multiple comparisons (*** $P < 0.001$). Sample sizes: *Tissue-enriched*: FG = 410, Testis = 1,121, Liver = 536, Ubiquitous = 344; *Tissue-restricted*: FG = 49, Testis = 399, Liver = 107, Ubiquitous = 3,990.

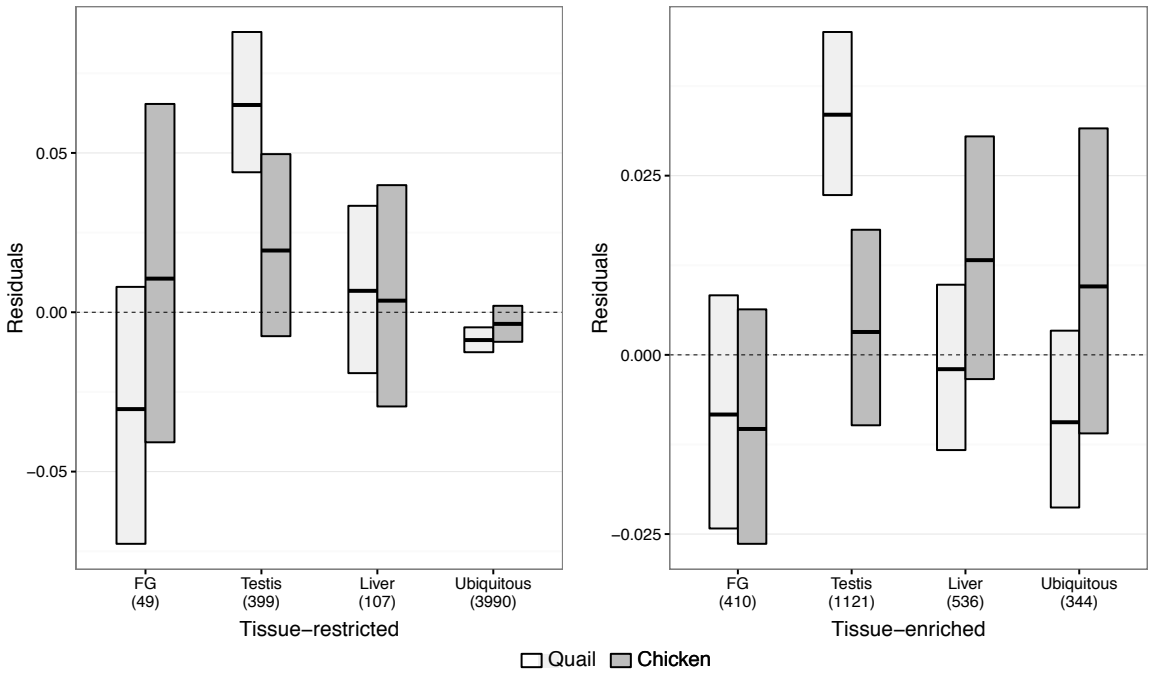

Figure 3.4. Average residuals (bar) and 95% CIs (boxes) from models regressing ω estimates along a single lineage (quail or chicken) on ω estimated from the rest of the tree. Estimates are based on 1:1:1:1 orthologs between quail, chicken, turkey, and zebra finch. Lineage-specific estimates from quail are black and those from chicken are grey. Genes that were enriched in a particular tissue are in the left panel, and those that were restricted to a tissue are in the right panel. Tissue-specific classifications were based on expression in quail as in Figure 3.1. The dotted line indicates values of 0. CIs derived from bootstrap resamplings of the mean without assuming normality. Sample sizes are in parentheses. FG = foam gland.

Although genes that encode male reproductive proteins often exhibit rapid evolution, heterogeneity exists in evolutionary rates across groups of genes expressed in particular tissues (Dean *et al.*, 2008; Grassa & Kulathinal, 2011), at different times during development (Good, 2005), with varying degrees of tissue- (Good, 2005; Dean *et al.*, 2008; Grassa & Kulathinal, 2011) or species-(Marshall *et al.*, 2010; Grassa & Kulathinal, 2011), specificity and from different functional classes (Dorus *et al.*, 2006; Turner *et al.*, 2008; Dorus *et al.*, 2010; Good *et al.*, 2013). Our data also suggest heterogeneous patterns of molecular evolution, as we find contrasting evolutionary rates for genes derived from two different male avian reproductive tissues. While genes from the foam gland diverge slowly, those from the testis evolve quite rapidly (Table 3.1, Figure 3.2). This pattern is not driven by differences in d_S across tissue classes and is robust to variation in expression level (Figure 3.2,S5).

We were also able to conclude that the genes encoding the reproductive function of testis and foam gland were responsible for their contrasting rates of protein evolution. By manipulating photoperiod, we produced males with reproductive organs in their breeding (*i.e.,* sexually active) or winter (*i.e.*, inactive) conditions. This allowed us to identify genes that are specifically up-regulated when the foam gland or testis actively make foam or sperm. Up-regulated genes evolve significantly faster if they derive from the testis rather than the foam gland (Table 3.1). However, genes not up-regulated when testes or foam glands are in breeding condition evolve at similar rates across the two tissues. This suggests that the heterogeneity in evolutionary rate is produced by the subset of genes that are actively involved in the production of foam and sperm.

Heterogeneity in evolutionary rates has been hypothesized to be a byproduct of compartmentalization of adaptations in response to sexual selection (Dorus *et al.*, 2006; Dean *et al.*, 2009; Dorus *et al.*, 2010). The logic here is that proteins with the potential to interact with the environment (either the male or female reproductive tract or other sperm) are expected to evolve more rapidly than those that do not. Consistent with this, proteins from mouse seminal vesicles evolve rapidly and produce the copulatory plug, which may influence the outcome of sperm competition or sexual conflict (Dean *et al.*, 2009; 2011; Dean, 2013). Likewise, sperm cell membrane and acrosome genes evolve faster relative to other sperm proteins in the mouse (Dorus *et al.*, 2010). The fact that foam gland genes evolve slowly would appear to be inconsistent with this hypothesis, because the entire foam complement is passed on to females and interacts with sperm and possibly the female reproductive tract to mediate the outcome of sperm competition (Klemm *et al.*, 1973, Finseth *et al*., 2013b, unpublished data). However, many of these interactions may not be antagonistic and antagonistic coevolution generally facilitates rapid evolution (see Discussion: *Morphological and functional divergence without sequence divergence*). Moreover, our analyses include all genes enriched in the foam gland or up-regulated when making foam, not just the subset that interacts with the environment (*i.e.*, those encoding foam proteins). Studies that distinguish genes that encode foam proteins from the rest of the genes expressed in the foam gland will allow a more rigorous test of the "compartmentalization" hypothesis.

Testis-expressed proteins evolve rapidly

Our work expands on previous research suggesting that the observation of rapidly evolving reproductive proteins extends to avian taxa. In most species, females evolve mechanisms to avoid polyspermy, as multiple sperm entering the egg results in embryo mortality. In the avian fertilization system, there is no selective disadvantage to polyspermy, as females tolerate physiological polyspermy whereby multiple sperm always enter the egg (Perry, 1987; Snook *et al.*, 2011). This is of interest because sexual conflict over polyspermy avoidance has been cited as a potential driver of the rapid evolution of reproductive proteins (Swanson & Vacquier, 2002a,b). Earlier studies in avian taxa showed that *i*) a few individual gamete-recognition genes evolve adaptively, while others do not (Berlin & Smith, 2005; Calkins *et al.*, 2007; Berlin *et al.*, 2008) and *ii*) genes expressed in chicken testis have lower than expected rates of nonsynonymous change (Grassa & Kulathinal, 2011), but those specifically expressed in the testis (N =12) exhibit elevated rates of nonsynonymous change.

Despite physiological polyspermy in birds, we find strong evidence for the rapid evolution of testis genes. Testis-biased genes exhibit *i*) elevated rates of protein evolution (Table 3.1), *ii*) a greater incidence of positive selection ($\omega > 1$) than ubiquitously expressed genes (Table 3.1), *iii*) a burst of accelerated evolution along the quail lineage (Figure 3.3), and *iv*) relatively large accelerations in evolutionary rate (Figure 3.4). Therefore, sexual conflict over polyspermy avoidance is not the sole driver of the rapid evolution of reproductive proteins because birds tolerate polyspermy (Perry, 1987). Avian systems provide a rich source of comparative reproductive data for further work that could discriminate among other hypotheses, such as sperm competition, cryptic female choice, other types of sexual conflict, or host-pathogen avoidance. However, given that

heterogeneity of selective pressures appears to dominate male reproductive genes at both the class and individual level (present study, Good, 2005; Dorus *et al.*, 2006; Dean *et al.*, 2008; 2009; Dorus *et al.*, 2010; Grassa & Kulathinal, 2011; Good *et al.*, 2013), no single reason may be sufficient to explain the widespread observation of rapidly-evolving male reproductive genes.

The more a gene is enriched in the testis, the more rapidly it evolves, even after correcting for expression differences (Table 3.1: restricted > enriched, Figure 3.2). Similar patterns are found in genes specialized in the mouse epididymis (Dean *et al.*, 2008) or exclusively expressed in male reproductive tissues (Dean *et al.*, 2009). This is not surprising given that tissue-specific genes tend to evolve more rapidly than widely expressed genes (Duret & Mouchiroud, 2000; Zhang & Li, 2004; Liao, 2006; Larracuente *et al.*, 2008), yet foam gland and liver genes do not show similar trends (Table 3.1, Figure 3.2). In *Drosophila melanogaster*, sex-biased genes in sex-limited tissues evolve faster than other narrowly expressed genes (Meisel, 2011). While the testis is sex-limited, the foam gland is not, as females possess rudimentary foam glands with limited secretory activity that do not make foam (McFarland *et al.*, 1968; Klemm *et al.*, 1973; King, 1981). Selection may therefore act more efficiently on testis-specific genes, as these are free from pleiotropic constraints associated with shared expression in female tissues like the foam gland.

Sperm flagella proteins evolve rapidly in quail

To understand what may be driving the rapid evolution of genes in quail, we focus on the function of genes with significant accelerations along the quail lineage and $\omega_{quail} > 1$, because these may have evolved under positive selection. (Functions are based on

annotations of the chicken orthologs.) Strikingly, four out of five genes with these characteristics are highly enriched in quail testis and also expressed (and sometimes restricted to) chicken testis (Table 3.1). One of this subset, *CASC1* (*Coja15527_c0_seq1*), is found in mouse whole sperm, while two others encode proteins related to axonemes, *RSPH1* (*Coja22788_c0_seq1*) and *DNAAF1* (*Coja19413_c0_seq1*) (Table 3.2). *CASC1* and *DNAAF1* (also known as *LRRC50* or *ODA7)* evolve rapidly in mice, with evidence for adaptive evolution of *DNAAF1* (Turner *et al.*, 2008; Dorus *et al.*, 2010).

Mutations to *DNAAF1* predispose zebrafish and humans to seminomas, a subclass of testicular germ cell tumors, possibly due to defects in normal germ cell regulation (Basten *et al.*, 2013). It is not surprising that two of the potentially adaptively evolving genes are both found in the testis and associated with cancer susceptibility (*DNAAF1* and *CASC1*). Cancer-related genes involved in tumor suppression and apoptosis are among those showing the strongest signatures of positive selection in primates (Nielsen *et al.*, 2005). Similar patterns have been hypothesized to be due to antagonistic pleiotropic effects of genes, implying a trade-off in reproductive benefits and disease susceptibility (Clark & Swanson, 2005). *DNAAF1* may be a prime candidate to explore the antagonistic pleiotropy hypothesis.

Variation in either of the axoneme-associated genes (*DNAAF1* and *RSPH1*) could have direct consequences for sperm performance and these genes may be targets of selection. Many sperm use flagella to propel themselves from the male through the female reproductive system towards the egg. Axonemes are the molecular motors of sperm flagella and cilia (Inaba, 2007; Satir & Christensen, 2007; Cummins, 2008; Inaba, 2011). The homolog of one gene evolving adaptively in quail, *RSPH1,* encodes proteins found in the

head of the radial spoke of *Chlamydomonas reinhardtii* axonemes (P. Yang, 2006). Its mouse homolog, *TSGA2*, encodes polypeptides localized to the sperm tail and anterior acrosome and maps to a locus associated with defects in sperm motility and sperm-egg interaction abnormalities (Hui *et al.*, 2006). Another gene likely under positive selection in quail, *DNAAF1,* codes for a cytoplasmic protein that regulates the assembly of dynein arms found in axonemes (Loges *et al.*, 2009; Duquesnoy *et al.*, 2009). Loss of function mutations to this gene in humans results in primary ciliary dyskinesia, a multifaceted disease of cilia one symptom of which is male infertility (Loges *et al.*, 2009; Duquesnoy *et al.*, 2009).

Although sperm flagella-associated genes tend to evolve slowly in other systems (Dorus *et al.*, 2010, but see Turner *et al.*, 2008), this class of genes may be a general target of selection due to sperm competition in birds. Rands *et al.,* (2013) scanned the *Geospiza magnirostris* genome for genes under positive selection along the passerine branch. Five of 47 genes they identified possessed ciliary related functions, some of which encode proteins known to be components of sperm flagella or with enriched expression in the testis. While the may be a general pattern in birds, it is notable that we observe accelerated and adaptive evolution of sperm flagella proteins along the *Coturnix* branch. Among Galliformes, Japanese quail sperm have several unique features including: *i*) flagella that are over twice as long as other non-passerine birds (Korn *et al.*, 2000), *ii*) exceptionally long midpieces covering 64-74% of total sperm flagellum length (Woolley, 1995; Korn *et al.*, 2000), and *iii*) midpieces that contain numerous mitochondria (1,400-2,500) per cell (Woolley, 1995; Korn *et al.*, 2000). In birds, sperm flagellum and midpiece length, as well as mitochondrial respiration, influence mobility, which is a heritable trait that potentially mediates PMSS (Froman & Feltmann, 1998; Birkhead *et al.*, 1999; Lupold *et al.*, 2009).

Taken together, the co-occurrence of quail-specific sperm attributes, rapidly and adaptively evolving sperm-flagella associated genes (Table 3.1), and accelerated evolution of testis-derived genes (Figures 3.3 and 3.4) suggest sperm experienced an unusual amount of change along the quail lineage when compared to other Galliformes. Alternatively, Japanese quail may experience more sperm competition than chickens, a species where dominant males mate guard in an effort to monopolize females (Cheng & Burns, 1988).

The dramatic sperm phenotypic evolution in *Coturnix* quail may not have been independent of foam. Previously, foam and sperm midpiece length have been hypothesized to coevolve. For example, sperm may require more energy (*i.e.*, mitochondria) to move through the foam or, alternatively, foam may provide materials such as lactate that are necessary for sperm motility (Korn *et al.*, 2000; Singh *et al.*, 2011b). Future investigations into the role of variation of rapidly evolving sperm flagella proteins in birds, as well as identification of protein constituents of foam, could provide original insights into aspects of sperm evolution.

Conclusions

The foam gland of male Japanese quail is a novel reproductive phenotype that influences the outcome of sperm competition. Although this accessory gland arose and changed dramatically in phenotype along the quail lineage, this was accomplished without concomitant major changes in protein-coding sequences. In fact, genes expressed in the foam gland actually evolve quite slowly, possibly due to a lack of involvement in antagonistic interactions or divergence in expression level. In contrast to the foam gland, genes with biased expression in the testis evolve rapidly and experienced a burst of

accelerated evolution along the quail lineage. Our data contributes to the growing body of evidence suggesting that the rapid evolution of reproductive tract genes may not be a universal byproduct of sexual selection. Rather, variation in selective pressure across tissues or individual genes likely dominated the evolution of reproductive proteins.

Acknowledgements

The authors would like to thank R. Meisel, B. Lazzaro, N. Chen, and the Harrison lab for helpful discussions and advice; E. Adkins-Regan for quail samples and housing; M. Mouchka, R. Bukowski and Q. Sun for assistance with transcriptome assembly and RNA-Seq data analysis; S. Bogdanowicz, J. Mosher, and P. Schweitzer for aiding with cDNA library preparation and Illumina sequencing; S. Iacovelli and N. Baran for quail handling; and T. van Deusen, S. Martin, L. Vann, and P. Smith for quail care. This research was funded by an NSF DDIG DEB-1010757 to FRF and RGH, P.E.O. Scholar, Cornell Sigma Xi, Andrew W. Mellon, and Paul. F. Feeny awards to FRF.

References

Adkins-Regan, E. 1999. Foam produced by male *Coturnix* quail: What is its function? *Auk* **116**: 184–193.

Almeida, F. & DeSalle, R. 2008. Evidence of adaptive evolution of accessory gland proteins in closely related species of the Drosophila repleta group. *Molecular Biology and Evolution* **25**: 2043–2053.

Andersen, C.L., Jensen, J.L. & Ørntoft, T.F. 2004. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer research* **64**: 5245–5250. AACR.

Anderson, M.J. & Dixson, A.F. 2002. Sperm competition: motility and the midpiece in primates. *Nature* **416**: 496–496. Nature Publishing Group.

Andersson, M.B. 1994. *Sexual selection*. Princeton University Press, Princeton, N.J.

Anisimova, M., Bielawski, J.P. & Yang, Z. 2002. Accuracy and power of Bayes prediction of amino acid sites under positive selection. *Molecular Biology and Evolution* **19**: 950–958.

Anisimova, M., Bielawski, J.P. & Yang, Z. 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Molecular Biology and Evolution* **18**: 1585–1592.

Arnqvist, G. 1998. Comparative evidence for the evolution of genitalia by sexual selection. *Nature* **393**: 784–786. Nature Publishing Group.

Basten, S.G., Davis, E.E., Gillis, A.J.M., van Rooijen, E., Stoop, H., Babala, N., *et al.,* 2013. Mutations in LRRC50 Predispose Zebrafish and Humans to Seminomas. *Plos Genetics* **9**: e1003384.

Benjamini, Y. & Hochberg, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological* **57**: 289–300.

Berlin, S. & Smith, N.G.C. 2005. Testing for adaptive evolution of the female reproductive protein ZPC in mammals, birds and fishes reveals problems with the M7-M8 likelihood ratio test. *BMC Evol Biol* **5**: 65.

Berlin, S., Qu, L. & Ellegren, H. 2008. Adaptive evolution of gamete-recognition proteins in birds. *J Mol Evol* **67**: 488–496.

Birkhead, T.R. & Fletcher, F. 1994. Sperm storage and the release of sperm from the sperm storage tubules in Japanese quail *Coturnix japonica*. *Ibis* **136**: 101–104.

Birkhead, T.R., Martinez, J.G., Burke, T. & Froman, D.P. 1999. Sperm mobility determines the outcome of sperm competition in the domestic fowl. *Proceedings Of The Royal Society B-Biological Sciences* **266**: 1759–1764.

Bork, P. & Koonin, E.V. 1998. Predicting functions from protein sequences--where are the bottlenecks? *Nat Genet* **18**: 313–318.

Calkins, J.D., El-Hinn, D. & Swanson, W.J. 2007. Adaptive evolution in an avian reproductive protein: zp3. *J Mol Evol* **65**: 555–563.

Carroll, SB. 2005. Evolution at two levels: On genes and form. *Plos Biol* **3**: 1159–1166.

Carroll, Sean B. 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**: 25–36.

Carroll, Sean, Grenier, J. & Weatherbee, S. 2001. *From DNA to diversity : molecular genetics and the evolution of animal design*.

Chen, H. & Boutros, P.C. 2011. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics* **12**: 35. BioMed Central Ltd.

Cheng, K.M. & Burns, J.R. 1998. Dominance relationship and mating behavior of domestic cocks- a model to study mate-guarding and sperm competition in birds. *Condor* **90**: 697 - 704.

Cheng, K.M., Hickman, A.R. & Nichols, C.R. 1989a. Role of the proctodeal gland foam of male Japanese quail in natural copulations. *Auk* **106**: 279–285.

Cheng, K.M., McIntyre, R.F. & Hickman, A.R. 1989b. Proctodeal gland foam enhances competitive fertilization in domestic Japanese quail. *Auk* **106**: 286–291.

Clark, N. & Swanson, W. 2005. Pervasive adaptive evolution in primate seminal proteins. *Plos Genetics* **1**: 335–342.

Clark, N.L., Aagaard, J.E. & Swanson, W.J. 2006. Evolution of reproductive proteins from animals and plants. *Journal of Reproduction and Fertility* **131**: 11–22.

Cummins, J. 2008. Sperm motility and energetics. In: *Sperm Biology: An evolutionary perspective* (T. R. Birkhead, D. Hosken, & S. Pitnick, eds). Academic Press, London.

Dean, M.D. 2013. Genetic disruption of the copulatory plug in mice leads to severely reduced fertility. *Plos Genetics* **9**: e1003185.

Dean, M.D., Clark, N.L., Findlay, G.D., Karn, R.C., Yi, X., Swanson, W.J., *et al.* 2009. Proteomics and comparative genomic investigations reveal heterogeneity in evolutionary rate of male reproductive proteins in mice (*Mus domesticus*). *Molecular Biology and Evolution* **26**: 1733–1743.

Dean, M.D., Findlay, G.D., Hoopmann, M.R., Wu, C.C., MacCoss, M.J., Swanson, W.J., *et al.* 2011. Identification of ejaculated proteins in the house mouse (*Mus domesticus*) via isotopic labeling. *BMC Genomics* **12**: 306. BioMed Central Ltd.

Dean, M.D., Good, J.M. & Nachman, M.W. 2008. Adaptive evolution of proteins secreted during sperm maturation: an analysis of the mouse epididymal transcriptome. *Molecular Biology and Evolution* **25**: 383–392.

Dorus, S., Busby, S.A., Gerike, U., Shabanowitz, J., Hunt, D.F. & Karr, T.L. 2006. Genomic and functional evolution of the Drosophila melanogaster sperm proteome. *Nat Genet* **38**: 1440–1445.

Dorus, S., Evans, P.D., Wyckoff, G.J., Choi, S.S. & Lahn, B.T. 2004. Rate of molecular evolution of the seminal protein gene SEMG2 correlates with levels of female promiscuity. *Nat Genet* **36**: 1326–1329.

Dorus, S., Wasbrough, E.R., Busby, J., Wilkin, E.C. & Karr, T.L. 2010. Sperm Proteomics Reveals Intensified Selection on Mouse Sperm Membrane and Acrosome Genes. *Molecular Biology and Evolution* **27**: 1235–1246.

Drummond, D.A., Bloom, J.D., Adami, C., Wilke, C.O. & Arnold, F.H. 2005. Why highly expressed proteins evolve slowly. *P Natl Acad Sci Usa* **102**: 14338–14343.

Duquesnoy, P., Escudier, E., Vincensini, L., Freshour, J., Bridoux, A.-M., Coste, A., *et al.* 2009. Loss-of-Function Mutations in the Human Ortholog of Chlamydomonas reinhardtii ODA7 Disrupt Dynein Arm Assembly and Cause Primary Ciliary Dyskinesia. *The American Journal of Human Genetics* **85**: 890–896. The American Society of Human Genetics.

Duret & Mouchiroud. 2000. Determinants of substitution rates in mammalian genes: Expression pattern affects selection intensity but not mutation rate. *Molecular Biology and Evolution* **17**: 68–74.

Eberhard, W. 1985. *Sexual selection and animal genitalia*.

Ellegren, H. & Parsch, J. 2007. The evolution of sex-biased genes and sex-biased gene expression. *Nature Reviews Genetics* **8**: 689–698.

Finn, S. & Civetta, A. 2010. Sexual Selection and the Molecular Evolution of ADAM Proteins. *J Mol Evol* **71**: 231–240.

Finseth, F.R. & Bondra, E.L., Harrison, R.G. 2013a. Combined proteomics and RNA-Seq reveal pervasive selective constraint during the evolution of a novel reproductive accessory fluid. Chapter 4, Ph.D. Thesis.

Finseth, F.R., Iacovelli, S.R., Harrison, R.G., Adkins-Regan, E.K. 2013b. A non-semen copulatory fluid influences the outcome of sperm competition in Japanese quail.

Journal of Evolutionary Biology, in press.

Finseth, F.R. & Harrison, R.G. 2013. A comparison of next generation sequencing technologies for transcriptome assembly and utility for RNA-Seq. Chapter 2, Ph.D. Thesis.

Froman, D.P. & Feltmann, A.J. 1998. Sperm mobility: a quantitative trait of the domestic fowl (Gallus domesticus). *Biol Reprod* **58**: 379–384.

Fujihara, N. 1992. Accessory reproductive fluids and organs in male domestic birds. *Worlds Poultry Science Journal* **48**: 39–56.

Good, J.M. 2005. Rates of Protein Evolution Are Positively Correlated with Developmental Timing of Expression During Mouse Spermatogenesis. *Molecular Biology and Evolution* **22**: 1044–1052.

Good, J.M., Wiebe, V., Albert, F.W., Burbano, H.A., Kircher, M., Green, R.E., *et al.* 2013. Comparative Population Genomics of the Ejaculate in Humans and the Great Apes. *Molecular Biology and Evolution*, doi: 10.1093/molbev/mst005.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., *et al.* 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* **29**: 644–652.

Grassa, C.J. & Kulathinal, R.J. 2011. Elevated Evolutionary Rates among Functionally Diverged Reproductive Genes across Deep Vertebrate Lineages. *International Journal of Evolutionary Biology* **2011**: 1509–1514.

Herlyn, H. & Zischler, H. 2007. Sequence evolution of the sperm ligand zonadhesin correlates negatively with body weight dimorphism in primates. *Evolution* **61**: 289–298.

Hornett, E.A. & Wheat, C.W. 2012. Quantitative RNA-Seq analysis in non-model species: assessing transcriptome assemblies as a scaffold and the utility of evolutionary divergent genomic reference species. *BMC Genomics* **13**: 1–1. BMC Genomics.

Hosken, D.J. 1998. Testes mass in megachiropteran bats varies in accordance with sperm competition theory. *Behav Ecol Sociobiol* **44**: 169–177. Springer.

Hui, L., Lu, J., Han, Y. & Pilder, S.H. 2006. The mouse T complex gene Tsga2, encoding polypeptides located in the sperm tail and anterior acrosome, maps to a locus associated with sperm motility and sperm-egg interaction abnormalities. *Biol Reprod* **74**: 633–643.

Hurle, B., Swanson, W., NISC Comparative Sequencing Program & Green, E.D. 2007. Comparative sequence analyses reveal rapid and divergent evolutionary changes of the WFDC locus in the primate lineage. *Genome Research* **17**: 276–286.

Inaba, K. 2007. Molecular Basis of Sperm Flagellar Axonemes: Structural and Evolutionary Aspects. *Annals of the New York Academy of Sciences* **1101**: 506–526.

Inaba, K. 2011. Sperm flagella: comparative and phylogenetic perspectives of protein components. *Molecular Human Reproduction* **17**: 524–538.

Kelleher, E.S., Swanson, W.J. & Markow, T.A. 2007. Gene Duplication and Adaptive Evolution of Digestive Proteases in Drosophila arizonae Female Reproductive Tracts. *Plos Genetics* **3**: e148.

Khaitovich, P. 2005. Parallel Patterns of Evolution in the Genomes and Transcriptomes of Humans and Chimpanzees. *Science* **309**: 1850–1854.

King, A.S. 1981. Cloaca. In: *Form and function in birds* (A. S. King & J. McLelland, eds), pp. 63–105. Academic Press, New York.

Klemm, R., Knight, C.E. & Stein, S. 1973. Gross and microscopic morphology of glandula-proctodealis (foam gland) of *Coturnix c. japonica* (aves). *J Morphol* **141**: 171–184.

Koonin, E.V. 2005. Orthologs, paralogs, and evolutionary genomics. *Annual Review of Genetics* **39**: 309–338.

Korn, N., Thurston, R., Pooser, B. & Scott, T. 2000. Ultrastructure of spermatozoa from Japanese quail. *Poultry Sci* **79**: 407–414.

Kumar, S. & Blaxter, M.L. 2010. Comparing de novo assemblers for 454 transcriptome data. *BMC Genomics* **11**: 571.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., *et al.* 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947–2948.

Larracuente, A.M., Sackton, T.B., Greenberg, A.J., Wong, A., Singh, N.D., Sturgill, D., *et al.* 2008. Evolution of protein-coding genes in Drosophila. *TRENDS in Genetics* **24**: 114–123.

Lemos, B., Bettencourt, B.R., Meiklejohn, C.D. & Hartl, D.L. 2005. Evolution of proteins and gene expression levels are coupled in Drosophila and are independently associated with mRNA abundance, protein length, and number of protein-protein interactions. *Molecular Biology and Evolution* **22**: 1345–1354.

Li, H. & Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., *et al.* 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Liao, B.Y. 2006. Impacts of Gene Essentiality, Expression Pattern, and Gene Compactness on the Evolutionary Rate of Mammalian Proteins. *Molecular Biology and Evolution* **23**:

2072–2080.

Loges, N.T., Olbrich, H., Becker-Heck, A., HAffner, K., Heer, A., Reinhard, C., *et al.* 2009. Deletions and Point Mutations of LRRC50 Cause Primary Ciliary Dyskinesia Due to Dynein Arm Defects. *The American Journal of Human Genetics* **85**: 883–889. The American Society of Human Genetics.

Long, M., Betrán, E., Thornton, K. & Wang, W. 2003. The origin of new genes: glimpses from the young and old. *Nature Reviews Genetics* **4**: 865–875.

Lupold, S., Calhim, S., Immler, S. & Birkhead, T.R. 2009. Sperm morphology and sperm velocity in passerine birds. *Proceedings Of The Royal Society B-Biological Sciences* **276**: 1175–1181.

Marshall, J.L., Huestis, D.L., Garcia, C., Hiromasa, Y., Wheeler, S., Noh, S., *et al.* 2010. Comparative Proteomics Uncovers the Signature of Natural Selection Acting on the Ejaculate Proteomes of Two Cricket Species Isolated by Postmating, Prezygotic Phenotypes. *Molecular Biology and Evolution* **28**: 423–435.

Martin-Coello, J., Dopazo, H., Arbiza, L., Ausio, J., Roldan, E.R.S. & Gomendio, M. 2009. Sexual selection drives weak positive selection in protamine genes and high promoter divergence, enhancing sperm competitiveness. *Proceedings Of The Royal Society B-Biological Sciences* **276**: 2427–2436.

McCarthy, D.J., Chen, Y. & Smyth, G.K. 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research* **40**: 4288–4297.

McFarland, L. Z., Warner, R.L., Wilson, W.O. & Mather, F.B. 1968. Cloacal gland complex of japanese quail. *Experientia* **24**: 941–&.

Meisel, R.P. 2011. Towards a More Nuanced Understanding of the Relationship between Sex-Biased Gene Expression and Rates of Protein-Coding Sequence Evolution. *Molecular Biology and Evolution* **28**: 1893–1900.

Min, X.J., Butler, G., Storms, R. & Tsang, A. 2005. OrfPredictor: predicting protein-coding regions in EST-derived sequences. *Nucleic Acids Research* **33**: W677–W680.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. & Wold, B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Meth* **5**: 621–628.

Nichols, C.R. 1991. A comparison of the reproductive and behavioural differences of feral and domestic Japanese quail. University of British Columbia, masters thesis.

Nielsen, R., Bustamante, C., Clark, A.G., Glanowski, S., Sackton, T.B., Hubisz, M.J., *et al.* 2005. A Scan for Positively Selected Genes in the Genomes of Humans and Chimpanzees. *Plos Biol* **3**: e170.

Nuzhdin, S.V. 2004. Common Pattern of Evolution of Gene Expression Level and Protein Sequence in Drosophila. *Molecular Biology and Evolution* **21**: 1308–1317.

O'Connor, T.D. & Mundy, N.I. 2009. Genotype-phenotype associations: substitution models to detect evolutionary associations between phenotypic variables and genotypic evolutionary rate. *Bioinformatics* **25**: i94–i100.

O'Neil, S.T., Dzurisin, J.D., Carmichael, R.D., Lobo, N.F., Emrich, S.J. & Hellmann, J.J. 2010. Population-level transcriptome sequencing of nonmodel organisms Erynnis propertius and Papilio zelicaon. *BMC Genomics* **11**: 310.

Panhuis, T.M., Clark, N.L. & Swanson, W.J. 2006. Rapid evolution of reproductive proteins in abalone and Drosophila. *Philos T R Soc B* **361**: 261–268.

Pál, C., Papp, B. & Hurst, L.D. 2001. Highly expressed genes in yeast evolve slowly. *Genetics* **158**: 927–931.

Perry, M.M. 1987. Nuclear events from fertilisation to the early cleavage stages in the domestic fowl (Gallus domesticus). *J. Anat.* **150**: 99–109.

Ramm, S.A. & Stockley, P. 2010. Sperm competition and sperm length influence the rate of mammalian spermatogenesis. *Biology Letters* **6**: 219–221.

Ramm, S.A., Oliver, P., Ponting, C., Stockley, P. & Emes, R. 2008. Sexual selection and the adaptive evolution of mammalian ejaculate proteins. *Molecular Biology and Evolution* **25**: 207–219.

Rands, C.M., Darling, A., Fujita, M., Kong, L., Webster, M.T., Clabaut, C.L., *et al.* 2013. Insights into the evolution of Darwin's finches from comparative analysis of the Geospiza magnirostris genome sequence. *BMC Genomics* **14**: 1–1. BMC Genomics.

Robinson, M.D. & Oshlack, A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology* **11**: R25.

Robinson, M.D., McCarthy, D.J. & Smyth, G.K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.

Sachs, B.D. 1967. Photoperiodic control of cloacal gland of japanese quail. *Science* **157**: 201–&.

Satir, P. & Christensen, S.T. 2007. Overview of structure and function of mammalian cilia. *Annu. Rev. Physiol.* **69**: 377–400.

Schäfer, J. & Strimmer, K. n.d. A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics. *Statistical Applications in Genetics and Molecular Biology* **4**.

Schlupp, I., Riesch, R., Tobler, M., Plath, M., Parzefall, J. & Schartl, M. 2010. A novel, sexually selected trait in poeciliid fishes: female preference for mustache-like, rostral filaments in male Poecilia sphenops. *Behav Ecol Sociobiol* **64**: 1849–1855.

Schmittgen, T.D. & Livak, K.J. 2008. Analyzing real-time PCR data by the comparative CT method. *Nature Protocols* **3**: 1101–1108.

Seiwert, C. & Adkins-Regan, E. 1998. The foam production system of the male Japanese quail: Characterization of structure and function. *Brain Behav Evol* **52**: 61–80.

Singh, R.P., H Sastry, von, K., Shit, N., Pandey, N.K., Singh, K.B., Mohan, J., *et al.* 2011. Cloacal gland foam enhances motility and disaggregation of spermatozoa in Japanese quail (*Coturnix japonica*). *Theriogenology* **75**: 563–569. Elsevier Inc.

Singh, R.P., Sastry, K.V.H., Pandey, N.K., Shit, N., Agrawal, R., Singh, K.B., *et al.* 2011. Characterization of lactate dehydrogenase enzyme in seminal plasma of Japanese quail (Coturnix coturnix japonica). *THE* **75**: 555–562. Elsevier Inc.

Singh, R.P., Sastry, K.V.H., Pandey, N.K., Singh, K.B., Malecki, I.A., Farooq, U., *et al.* 2012. The role of the male cloacal gland in reproductive success in Japanese quail (*Coturnix japonica*). *Reprod. Fertil. Dev.* **24**: 405.

Singh, R.S. & Kulathinal, R. 2000. Sex gene pool evolution and speciation: A new paradigm. *Genes Genet Syst* **75**: 119–130.

Sittmann, K. & Abplanalp, H. 1965. Duration and recovery of fertility in Japanese quail (*Coturnix coturnix japonica*). *British Poultry Science* **6**: 245–250.

Snook, R.R., Hosken, D.J. & Karr, T.L. 2011. The biology and evolution of polyspermy: insights from cellular and functional studies of sperm and centrosomal behavior in the fertilized egg. *Reproduction* **142**: 779–792.

Swanson, W. & Vacquier, V. 2002a. Reproductive protein evolution. *Annual Review of Ecology and Systematics* **33**: 161–179.

Swanson, W. & Vacquier, V. 2002b. The rapid evolution of reproductive proteins. *Nature Reviews Genetics* **3**: 137–144.

Tatusov, R.L. 1997. A Genomic Perspective on Protein Families. *Science* **278**: 631–637.

True, J. & Carroll, S. 2002. Gene co-option in physiological and morphological evolution. *Annual Review of Cell and Developmental Biology* **18**: 53–80.

Turner, L., Chuong, E. & Hoekstra, H. 2008. Comparative analysis of testis protein evolution in rodents. *Genetics* **179**: 2075–2089.

Turner, L.M. & Hoekstra, H.E. 2008. Causes and consequences of the evolution of reproductive proteins. *Int. J. Dev. Biol.* **52**: 769–780.

van Tuinen, M. & Dyke, G. 2004. Calibration of galliform molecular clocks using multiple fossils and genetic partitions. *Molecular Phylogenetics And Evolution* **30**: 74–86.

Wagner, G. & Lynch, V. 2005. Molecular evolution of evolutionary novelties: The vagina and uterus of therian mammals. pp. 580–592.

Wagstaff, B.J. 2005. Molecular population genetics of accessory gland protein genes and testis-expressed genes in *Drosophila mojavensis* and *D. arizonae*. *Genetics* **171**: 1083–1101.

Walters, J.R. & Harrison, R.G. 2011. Decoupling of rapid and adaptive evolution among seminal fluid proteins in *Heliconius* butterflies with divergent mating systems. *Evolution* **65**: 2855–2871.

Wang, Z., Gerstein, M. & Snyder, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10**: 57–63.

Wiens, J.J., Sparreboom, M. & Arntzen, J.W. 2011. Crest evolution in newts: implications for reconstruction methods, sexual selection, phenotypic plasticity and the origin of novelties. *J Evolution Biol* **24**: 2073–2086.

Williford, A., Stay, B. & Bhattacharya, D. 2004. Evolution of a novel function: nutritive milk in the viviparous cockroach, Diploptera punctata. *Evol Dev* **6**: 67–77.

Wong, A. 2010. Testing the effects of mating system variation on rates of molecular evolution in primates. *Evolution* **64**: 2779–2785.

Wong, A. 2011. The molecular evolution of animal reproductive tract proteins: what have we learned from mating-system comparisons? *International Journal of Evolutionary Biology* **2011**: 1–9.

Woolley, D.M. 1995. The structure of the spermatozoon of the Japanese quail, Coturnix coturnix L., var. japonica. *Acta Zoologica* **76**: 45–50. Wiley Online Library.

Wray, G.A. 2007. The evolutionary significance of cis-regulatory mutations. *Nature Reviews Genetics* **8**: 206–216.

Yang, P. 2006. Radial spoke proteins of Chlamydomonas flagella. *Journal of Cell Science* **119**: 1165–1174.

Yang, Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution* **24**: 1586–1591.

Yang, Z., Nielsen, R., Goldman, N. & Pedersen, A. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**: 431–449.

Zhang, L. & Li, W.-H. 2004. Mammalian housekeeping genes evolve more slowly than tissue-specific genes. *Molecular Biology and Evolution* **21**: 236–239.

Zhang, Z., Li, J., Zhao, X.-Q., Wang, J., Wong, G.K.-S. & Yu, J. 2006. KaKs_Calculator: Calculating Ka and Ks Through Model Selection and Model Averaging. *Genomics, Proteomics & Bioinformatics* **4**: 259–263.

Zhang, Z., Xiao, J., Wu, J., Zhang, H., Liu, G., Wang, X., *et al.* 2012. ParaAT: A parallel tool for constructing multiple protein-coding DNA alignments. *Biochemical and Biophysical Research Communications* **419**: 779–781. Elsevier Inc.

Zhao, S. & Fernald, R.D. 2005. Comprehensive algorithm for quantitative real-time polymerase chain reaction. *Journal of Computational Biology* **12**: 1047–1064. Mary Ann Liebert, Inc. 2 Madison Avenue Larchmont, NY 10538 USA.

CHAPTER 4

COMBINED PROTEOMICS AND RNA-SEQ REVEAL CONTRASTING PATTERNS OF CONSTRAINT AND DIVERGENCE DURING THE EVOLUTION OF A NOVEL REPRODUCTIVE FLUID

Abstract

Proteins that males transfer to females in seminal fluid are important determinants of reproductive fitness. Here, we focus on an example of a novel seminal fluid-like proteome and characterize its protein constituents. Male Japanese quail possess a novel 'foam gland' that secretes a viscous fluid that is whipped into a stiff, meringue-like foam. During mating, males transfer foam to females. Foam enhances male reproductive success through improving sperm performance and mediating sperm competition. We manipulated the foam glands of males and used RNA-Seq to identify genes that increased in expression as foam glands transformed from an inactive to an active state. We compared this dataset with putative foam proteins identified via standard proteomics techniques to generate a list of high confidence foam proteins. We report contrasting patterns of conservation and divergence in foam proteins. The foam proteome disproportionately utilizes genes with evolutionarily ancient origins that are under selective constraint. Conversely, dominant foam proteins are novel and secreted foam proteins diverge rapidly. Our proteomic data suggests that foam may benefit male reproductive fitness in diverse ways, which could explain the contrasting evolutionary scenarios. Foam may function in part through emergent properties of co-opted ancestral genes to facilitate aerobic respiration of sperm, though individual foam proteins may contribute to sperm defense and could potentially evolve rapidly due to antagonistic coevolution. Thus, novel proteomes and functions can arise from both re-purposing of conserved proteins in new combinations and rapid divergence of individual proteins.

Introduction

In internally fertilizing animals, sperm is transferred to females together with a cocktail of proteins secreted by specialized accessory organs. These seminal fluid proteins (SFPs) have profound effects on reproductive success for both sexes. SFPs play key roles in male fertility through influencing the nourishment, protection, motility and storage of sperm (Poiani, 2006). In females, receipt of SFPs can be beneficial, *e.g.*, supplying food or antibiotic properties, or costly, *e.g.,* suppressing the female immune system or inhibiting re-mating (James & Hargreave, 1984; Poiani, 2006; Ram & Wolfner, 2007). SFPs are also major players in post-mating sexual selection and mediate the outcome of sperm competition or intersexual conflict (Clark *et al.*, 1995; Chapman *et al.*, 1995; Stockley, 1997; Poiani, 2006; Ram & Wolfner, 2007). Removal of SFPs underscores their significance in reproduction, as males without certain SFPs suffer major fertility defects (*e.g.,* Dean, 2013), or lose in sperm competition (*e.g.,* Finseth *et al.* 2013b, Cheng *et al.*, 1989a; Chapman *et al.*, 2000).

Despite the fundamental importance of ejaculate-female interactions to successful fertilization, SFPs exhibit rapid evolutionary divergence (Swanson & Vacquier, 2002; Clark *et al.*, 2006; L. M. Turner & Hoekstra, 2008). SFPs interact with sperm proteins, SFPs from other males, the female reproductive tract, and pathogens (Poiani, 2006; Pitnick *et al.*, 2008). These interactions can be antagonistic or cooperative, setting the stage for coevolution to occur. Coevolutionary dynamics may drive the rapid divergence of SFPs and can result in reproductive isolation between species (Swanson & Vacquier, 2002; Coyne & Orr, 2004; Rice, 1998).

Because of their critical role in reproduction and sexual selection, the functional and evolutionary significance of SFPs have been studied extensively. One common approach is

to use a combination of proteomic and genomic techniques to characterize the collection of SFPs transferred to females during mating (Andrés *et al.*, 2008; Findlay *et al.*, 2009; Ramm *et al.*, 2009; Dean *et al.*, 2009; Kelleher *et al.*, 2009; Baer *et al.*, 2009; Findlay & Swanson, 2010; Walters & Harrison, 2010; Dean *et al.*, 2011). This line of research has provided important insights into the complexity and functional role of seminal fluid, but thus far has been restricted to studies of insects and mammals (Poiani, 2006). SFPs are taxonomically widespread and their roles may vary in taxon-specific ways. For example, seminal fluid from species with a cloaca, which serves as a common opening for the reproductive and excretory systems, may be under additional selective pressure to provide antimicrobial substances or elicit a female immune response.

Foam produced by male Japanese quail (*Coturnix japonica*) is one example of the remarkable diversity of animal accessory gland fluids. Male Japanese quail possess a specialized accessory organ (the 'foam gland') that secretes a viscous protein complex (Coil & Wetherbee, 1959; McFarland, L. Z. *et al.*, 1968; Klemm *et al.*, 1973; King, 1981). Upon encountering a female, males use their cloacal sphincter muscle to rhythmically whip the secretion into a stiff, meringue-like foam (Figure 4.1; Seiwert & Adkins-Regan, 1998). During mating, males transfer semen, along with copious amounts of foam, to the female reproductive tract. Foam is produced separately from the seminal fluid and sperm, and semen and foam are not mixed until they are inside the female (Ikeda & Taji, 1954; King, 1981). Thus, foam is not a seminal fluid *per se,* as it is never packaged with sperm in semen. However, foam and seminal fluid likely play similar functional roles. The presence of foam mediates the outcome of sperm competition (Finseth *et al.* 2013b, Cheng *et al.*, 1989a; Adkins-Regan, 1999), increases fertilizing efficiency at certain stages of ovulation

(Adkins-Regan, 1999), improves sperm motility and viability (Cheng *et al.*, 1989b; Singh *et al.*, 2011a), enhances sperm transport in the oviduct (Singh *et al.*, 2012), extends the window of time a male can fertilize any eggs (Cheng *et al.*, 1989a; Singh *et al.*, 2012), and disaggregates clumped sperm (Singh *et al.*, 2011a).

In this investigation, we characterize the protein constituents of foam. In addition to being the first broad survey of an accessory gland proteome in birds, our study offers insight into the evolutionary processes that shape novel copulatory fluids. Foam is unique to *Coturnix* males; no other group, including sister genera, possess a secretion with a similar foamy quality (Klemm *et al.*, 1973; Fujihara, 1992). Indeed, most birds have no dedicated accessory organs and any seminal fluid is produced from the seminiferous tubules or epithelia of the testis or ductus deferens (Lake, 1981).

We used a standard tandem mass-spectrometry (MS/MS) proteomic approach to identify a list of putative foam proteins. Because the foam gland is closely associated with the cloaca, complete separation of foam from potential cloacal contaminants is not possible. Therefore, we employed deep sequencing of RNA ('RNA-Seq') to identify the subset of proteins with increased expression at the onset of sexual activity; these genes are likely to encode foam proteins. We took advantage of the fact that the foam gland is androgen-dependent and manipulated photoperiod and testosterone levels to produce males that did or did not make foam (Figure 4.1). The genes significantly up-regulated in the foam glands of males actively producing foam *and* identified by proteomics were considered to be the best candidates to encode foam proteins. We used chicken (*Gallus gallus*) orthologs to annotate the complete set of putative foam proteins. Our integrative, comparative approach allowed us to examine: 1) the functions of major foam gland proteins and how

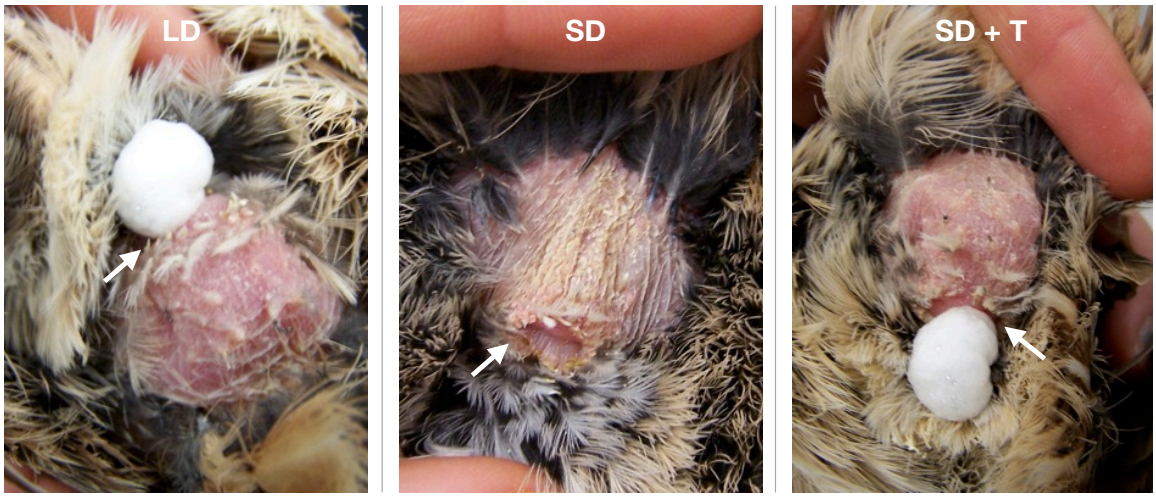

Figure 4.1 Foam glands and foam production respond to different photoperiod and hormone treatments. Foam glands and foam of males from three different photoperiod and hormone treatments (N = 6 per treatment). Photoperiod treatments mimicked breeding (*i.e.*, long day) or winter (*i.e.*, short day) conditions. All males were implanted with either an empty implant or one filled with testosterone. All foam glands were gently squeezed prior to taking the picture. Treatments are as follows: LD = long day males, foam actively produced; SD = short day males, note the regression of the foam gland and lack of foam; SD + T = short day males plus testosterone, foam actively produced. Arrow indicates cloacal vent. Photos by F. Finseth and S. Iacovelli.

they compare with SFPs from other taxa, 2) the origins of foam genes and 3) the selective pressures that shape foam protein evolution. We discuss our results in light of the general evolutionary significance of reproductive proteins, as well as in the context of the specialized reproductive biology of *C. japonica.*

Methods

Subjects

Unless stated otherwise, Japanese quail were lab-reared and housed in a 16L:8D light:dark cycle. Birds were housed individually at four weeks of age (the onset of sexual maturity is six weeks of age). Males were pre-screened for mating competency and only those males who successfully mated with a female at least once were included in the study. Prior to the start of the experiment, all males were weighed and their tarsus lengths and foam gland areas (length x width) were measured. Males were distributed randomly among treatment groups according to mass, mass/tarsus length (a proxy for condition), and foam gland area/mass. All animal procedures were approved by Cornell University IACUC.

Gene expression analyses ('GE'*)*

Experimental design and methods

Two groups of adult Japanese quail males were used to identify genes encoding foam proteins through differential gene expression ('GE') analysis. Birds in Group 1 were twelve months old at the start of the experiment (sample size ('N') = 6, experiments ran

from November 2010 - February 2011). Birds in Group 2 were two months old at the start of the experiment (N = 12, experiments ran from August - October 2011). Each group was distributed equally into three treatments: long day ('LD'), short day ('SD'), and short day + testosterone ('SD + T') for a total of six males per treatment. LD and SD+T males have functional foam glands and produce foam, whereas SD males have regressed foam glands that do not make foam (Figure 4.1). Initially, all males were housed in a 16L:8D light:dark cycle to simulate breeding conditions. The SD and SD + T males were later placed in a short day 8L:16D light:dark cycle for either 3 (Group 2) or 7 (Group 1) weeks prior to hormone implantation (below). The LD males continued on long days for the same amount of time. Short day treatments cause both the foam gland and testes to regress and cease producing foam and sperm, respectively (Sachs, 1967). All males were surgically implanted with either empty (LD, SD) or testosterone-filled (SD + T; Sigma Aldrich®) implants. Testosterone implants produce rapid development of the foam gland in short day males (Schumacher & Balthazart, 1983). Thus, the SD + T treatment allowed us to control for gene expression differences that are determined by photoperiod, but not important for foam production.

Two Silastic™ implants (25 mm length, 1.6 mm inner diameter, and 2.4 mm outer diameter) of the appropriate treatment were placed subcutaneously in the neck/upper back region after numbing the skin with Bupivacaine (Sigma-Aldrich®). The incision site was closed with 1-2 stitches and sealed with VetBond. Implants were checked the following day for proper insertion and any remaining sutures were removed after one week. Throughout the experiment, foam gland area, production, and volume (Group 2 only) were monitored on a weekly basis. Foam production was classified as either "Yes"

(producing any foam) or "No" (no foam produced). We report measurements at three time points: 1) prior to light treatments (*i.e.*, baseline), 2) immediately preceding implantation (*i.e.*, after photoperiod treatment), and 3) ~ five weeks after implantation (*i.e.*, after hormone treatment). We confirmed that all males exhibited the foam gland phenotype appropriate to their treatment (LD: large foam gland, producing foam; SD: regressed foam gland, not producing foam; SD + T: large foam gland, producing foam) (Figures 4.1 & 4.2, Table S4.2). We euthanized subjects approximately five weeks after implantation and immediately dissected out their foam glands. Foam gland samples were immediately frozen in liquid nitrogen, and later moved to -80 °C until RNA extraction.

Library preparation, sequencing, and differential gene expression analysis

We extracted RNA from 18 foam glands (6 males x 3 treatments) with the Agencourt® RNAdvance™ Tissue Kit (Beckman Coulter) following the manufacturer's instructions except that we used half-reactions. RNA quality and concentration were assessed by agarose gel electrophoresis and NanoDrop™ spectrophotometry. We confirmed RNA purity and integrity using an Agilent 2100 BioAnalyzer. In January 2012, we prepared 18 foam gland libraries from 1.2 µg total RNA using the TruSeq™ RNA Sample Preparation Kit (Illumina®) following the manufacturer's instructions. All samples were tagged with a unique adapter index, pooled, and single-end sequenced on two lanes of an Illumina HiSeq™ 2000, with a target read length of 100 bp. Sequencing was performed by the Cornell University Life Sciences Core Laboratories Center in April 2012.

Initial quality filtering and barcode removal were performed by the Genomics Facility at Cornell University's Institute of Biotechnology. Unless stated otherwise,

bioinformatics analyses were conducted on a Linux, Dell Precision T3500n with 4 cores, 24 GB RAM, and 4 TB HDD, housed at the Cornell Biology Service Unit of Cornell University's Institute of Biotechnology. We used fastq-mcf (http://code.google.com/p/ea-utils/wiki/FastqMcf) to remove Illumina adaptors, trim low-quality terminal ends, discard short sequences, and filter reads with phred scores < 20. In an earlier study, we constructed two versions of transcriptomes from foam gland, liver, and testis tissues from six Japanese quail males (Finseth *et al.,* 2013a). One version represented an initial, 'exhaustive' transcriptome (N = 81,868 transcripts). The second version was a subset of the first consisting of high-quality contigs that we believed represented real genes, because they were expressed above some threshold level ('filtered transcriptome'; N = 24,035 transcripts). Filtered reads from each foam gland sample were aligned to the exhaustive transcriptome using the Burrow-Wheeler transform in BWA (Li & Durbin, 2009). The number of reads per sample uniquely mapped to each transcript was tabulated with samtools as implemented in custom python scripts (Li *et al.*, 2009).

Our goal was to identify transcripts up-regulated when the foam gland was actively making foam. To this end, we tested for differential expression of transcripts in LD or SD + T versus SD treatments using the multifactor glm approach in EdgeR version 3.2.3 (Robinson *et al.*, 2010). Briefly, we filtered out very lowly expressed transcripts and normalized our RNA-Seq data using the TMM approach (Robinson & Oshlack, 2010). We fit negative binomial GLMs with Cox-Reid tagwise dispersion estimates to models that incorporated photoperiod and hormone treatment as factors. Our design matrix specified contrasts to find genes differentially expressed in 1) LD versus SD and 2) SD + T versus SD. We used the false discovery rate to account for multiple testing and applied a cutoff of 0.05

to consider a transcript 'significantly' differentially expressed in a particular comparison (Benjamini & Hochberg, 1995). Transcripts that were 1) significantly up-regulated in LD relative to SD, 2) significantly up-regulated in SD + T relative to SD, and 3) found in the 'filtered transcriptome' comprised a list of transcripts that may encode foam proteins (N = 2,676). The candidate genes meeting this set of criteria are referred to as '*GE*' because they were identified by differential *g*ene *e*xpression analyses.

*Proteomics analyses ('*P'*)*

We used MS/MS to identify the protein constituents of foam. Details regarding protein preparation, mass spectrometry, and bioinformatics identification of peptides are included in the supplementary methods (Methods S4.1). In brief, we pooled foam from six males, purified it, and ran this on a 1D SDS PAGE gel for protein separation. Gel slices were digested with trypsin into peptides. The resulting peptides were extracted and fractionated using nano liquid chromatography prior to two rounds of mass spectrometry (nanoLC-MS/MS). Spectra were searched against the predicted open reading frames of the 'exhaustive' transcriptome (Finseth *et al*, 2013a). Matches at or above the 99% confidence threshold were considered confidently matched peptides. Proteins with at least 2 unique peptide matches constituted a preliminary list of 1006 potential genes encoding foam proteins ('*P*'). This was compared to the *GE* list of 2,676 transcripts identified as significantly up-regulated when foam is produced. The overlapping list of 253 transcripts we consider 'high confidence' foam proteins ('*P + GE*').

Protein abundance and bioinformatics analyses

Protein abundance was calculated with the exponentially modified protein abundance index (emPAI) statistic (Ishihama *et al.*, 2005). For each transcript, expression levels were estimated as reads per kilobase per million mapped reads (RPKM; Mortazavi *et al.*, 2008). We regressed emPAI on average RPKMs when foam was active (LD and SD + T treatment groups). The data were log-transformed prior to regression (excluding proteins with RPKMS = 0).

We annotated quail transcripts by identifying orthologous chicken sequences using the reciprocal best BLAST approach described in Finseth *et al.* (2013a). We annotated the *G. gallus* orthologs based on chicken proteins from Ensembl version 69 and linked Ensembl gene IDs with protein descriptions using the Biomart tool (http://www.biomart.org/). We removed common contaminants from the protein set (*i.e.*, keratin) unless they were also identified in the *GE* list for a total of 1002 *P* proteins. The *P* proteins were annotated with respect to eight criteria: 1) identification in the protein or gene expression analyses, 2) fold change in foam inactive versus active glands (see Methods: *Gene expression analysis*), 3) rate of protein evolution (see Methods: *Rates of protein evolution*), 4) enriched or restricted gene expression in a tissue (Finseth *et al*., 2013a), 5) the ortholog hit ratio (*i.e.*, the amino acid sequence length of the quail protein, standardized by the length of its chicken ortholog; O'Neil *et al.*, 2010; Van Belleghem *et al.*, 2012), 6) the presence of a signal peptide, 7) Gene Ontology (GO) biological process and molecular function terms, and 8) protein classes. We predicted the presence of a signal sequence in SignalP 4.1 using default parameters (Petersen *et al.*, 2011). GO terms and protein classes were obtained from the PANTHER database (http://www.pantherdb.org).

GO biological process and molecular function terms were clustered according to similarity using the DAVID tool version 6.7 (Huang *et al.*, 2008; 2009). Using a modified Fisher's exact test (*i.e.*, EASE score; Huang *et al.*, 2008), we tested for enrichment of clusters of terms in the genes encoding foam proteins (*P + GE, P*) compared with expectations from the filtered transcriptome. Enrichment scores above 1.3 were considered significant, because 1.3 is equivalent to a non-log scale value of 0.05 (Huang *et al.*, 2008).

Transcript categorizations

We categorized genes from the filtered transcriptome with a *G. gallus* ortholog two ways: according to their relative expression in the foam gland or to their spatial expression patterns across three tissues (categories described below). For the foam gland classes, genes were designated as either encoding a foam protein from both the proteomic and differential gene expression analyses (*P + GE*; N = 194), encoding a foam protein identified only proteomically (*P*; N = 466), significantly up-regulated in active foam glands (*GE*; N = 1,396), expressed in the foam gland (*FG exp*; N = 4652) or not expressed in the foam gland (*Other*: N = 2129). The *FG exp* class was determined in a previous study (Finseth *et al*. 2013a). We also categorized genes based on their spatial expression patterns according to the '*tissue-restricted*' categorization described in Finseth *et al*. (2013a). Genes were either: *P + GE* (N = 194), *P* (N = 466), only expressed in the foam gland (*FG*; N = 82), only expressed in the liver (*Liver*; N = 209), only expressed in the testis (*Testis*; N = 1003), or expressed in all three tissues (*Ubiquitous*; N = 5438). For both classifications, each transcript was represented in only the least inclusive group that was appropriate. For example, a

transcript that was both expressed in the foam gland (*FG Exp*) and up-regulated when the foam gland was active (*GE*) would have been classed as *GE*.

Rates of protein evolution

We assigned quail:chicken orthologs as described in Finseth *et al.* (2013a). Alignment of nucleotide sequences for each orthologous pair and estimation of pairwise rates of protein evolution (*i.e.*, ω, the ratio of nonsynonymous (d_N) to synonymous (d_S) substitution rates) are from (Finseth *et al*., 2013a). We identified 1:1:1:1 orthologs among quail:chicken:turkey:zebrafish, aligned them at the nucleotide level, and estimated whether there was a significant change in selective pressure (ω) along the branch leading to quail as described in Finseth *et al.* (2013a). We examined whether the proportion of genes with orthologs, the pairwise evolutionary rates, and the proportion of orthologs with significant shifts (accelerations or decelerations) in ω along the quail lineage varied according to foam gland or tissue classes (see Methods: *Transcript categorizations*). For the pairwise ω values, we constructed 95% confidence intervals (CIs) around mean estimates for each foam gland or tissue class. We also examined whether evolutionary rate varied according to whether or not a foam protein (*P*) was secreted as determined by the presence of a signal peptide. Since the signal peptide could have been truncated in genes that were only partially recovered in our transcriptome assembly, only those genes with ortholog hit ratios > 0.90 were used for the 'secreted' analysis.

Origin of foam genes

To identify the evolutionary origins of genes encoding foam proteins, we followed the general approach described by Knox and Baker (2008). First, we determined 1:1 single copy orthologs of transcripts from the exhaustive transcriptome with OrthoMCL (Chen *et al.*, 2006). OrthoMCL combines a reciprocal best BLAST approach with a graph-clustering algorithm to identify homologous proteins and distinguish potential orthologs from paralogs. We used the 'exhaustive' transcriptome in OrthoMCL, as we wanted to be confident that the best BLAST hit from chicken to quail was identified. However, we only have confidence that genes from the 'filtered' transcriptome are true genes and not assembly artifacts (described in Finseth *et al*., 2013a). Therefore, we restricted the list of confidently assigned orthologs to 1) the single best hit based on BLAST similarity scores, 2) those where the best hit was from *G. gallus,* and 3) those identified for transcripts from the filtered transcriptome characterized by Finseth *et al*. (2013a; N = 9, 774). Note that this criteria is slightly different than the one discussed under Methods: *Rates of protein evolution* and identifies slightly fewer orthologs (N = 10, 129). From this list, we identified the evolutionary origin of a quail transcript by finding the most distantly related species that possessed an orthologous gene (*i.e.*, had a member of the same orthologous group as identified by OrthoMCL). For each transcript, the most distantly related taxon with an ortholog was classed as Aves, Vertebrate, Animal, Eukaryote, or Bacteria + Archaea based on the appropriate least inclusive group. For example, if the most distantly related species with an ortholog for a particular transcript was *Mus musculus*, that transcript would have been categorized as having a 'Vertebrate' origin. We evaluated whether evolutionary origins of the transcripts varied according to either tissue-specific or foam gland expression categorizations as described under Methods: *Transcript categorizations*.

Real-time quantitative PCR validation of RNA-Seq

To validate RNA-Seq, we performed RT-qPCR on the 18 RNA samples used for RNA-Seq (6 replicates per treatment (SD, SD + T, LD)). RNA was treated with Turbo™ DNase (Ambion®) and confirmed to be free of genomic DNA by attempting to PCR amplify a panel of three housekeeping genes (see Table S4.1: *Treatment*) directly from RNA. Two hundred ng of RNA was reverse transcribed into cDNA with SuperScript® III First Strand cDNA Synthesis Kit (Invitrogen™) following the manufacturer's instructions. Primers were designed from nine transcripts significantly up-regulated when foam was active in both the SD + T and LD treatments. We verified that primers amplified the intended target by Sanger sequencing. β-actin served as an internal control and was confirmed to be stable across treatments using Normfinder (Adkins-Regan, 1999; Andersen *et al.*, 2004). RT-qPCR reactions (25 µL) were performed in duplicate with 33 ng of cDNA template and 200 nM of each primer using the Power SYBR® Green Master Mix (Applied Biosystems). Samples were run on a ViiA™ 7 (Applied Biosystems) thermocycler with the following parameters: 95° C for 10 min, 40 cycles of 95° C for 15s and 60° C for 60s. Primer efficiencies were calculated with Real-time PCRMiner and ranged from 95 - 100% (Zhao & Fernald, 2005). We calculated ΔCt and log fold change in the SD treatment versus one of the two 'foam active' treatments (SD + T or LD) as described by Zhao & Fernald (2005). We tested for correlations between log fold changes generated by RT-qPCR and RNA-Seq for both treatments separately. Primer sequences, target sizes, efficiencies, and log fold changes are given in Table S4.1.

Data analysis

Unless stated otherwise, all analyses were performed in R Version 3.0.1 (R Core Team 2013). CIs were generated in the Hmisc package of R (Harrell, 2012) by performing 10,000 bootstrap resamplings of the mean without assuming normality. Fisher's exact test was used to test for significant difference among proportions. To correct for multiple testing where necessary, the false discovery rate was applied with a cutoff of 0.05 to call significance, unless stated otherwise (Benjamini & Hochberg, 1995).

Results

Males responded to the light and hormone treatments as anticipated. The short day photoperiod treatment caused the SD and SD + T males' foam glands to regress and stop producing foam (Figures 4.1, 4.2; Table S4.2). Implanting testosterone caused the SD + T males' foam glands to recrudesce and produce normal volumes of foam (Figures 4.1, 4.2; Table S4.2). The RNA-Seq and qPCR data were highly correlated for SD and either LD or SD + T treatments (SD−LD: R^2 = 0. 9315, $F_{(1, 7)}$ = 95.16, P = 2.519 * 10^{-5}; SD−SD +T: R^2 = 0. 9624, $F_{(1,7)}$ = 179.1, P = 3.05 * 10^{-6}).

Protein identification, abundance, and enriched GO terms

Proteomic analyses identified 1002 proteins (*P*), 253 of which were also significantly up-regulated when foam was being made (*P + GE*; Table S4.3). Protein abundance (emPAI) and gene expression level (RPKM) were correlated for genes that encode foam proteins (Figure 4.3). The correlation is stronger for the subset of genes identified by both proteomics and differential gene expression analyses (*P + GE*), than for those identified only proteomically (*P*). Requiring that genes be both significantly up-

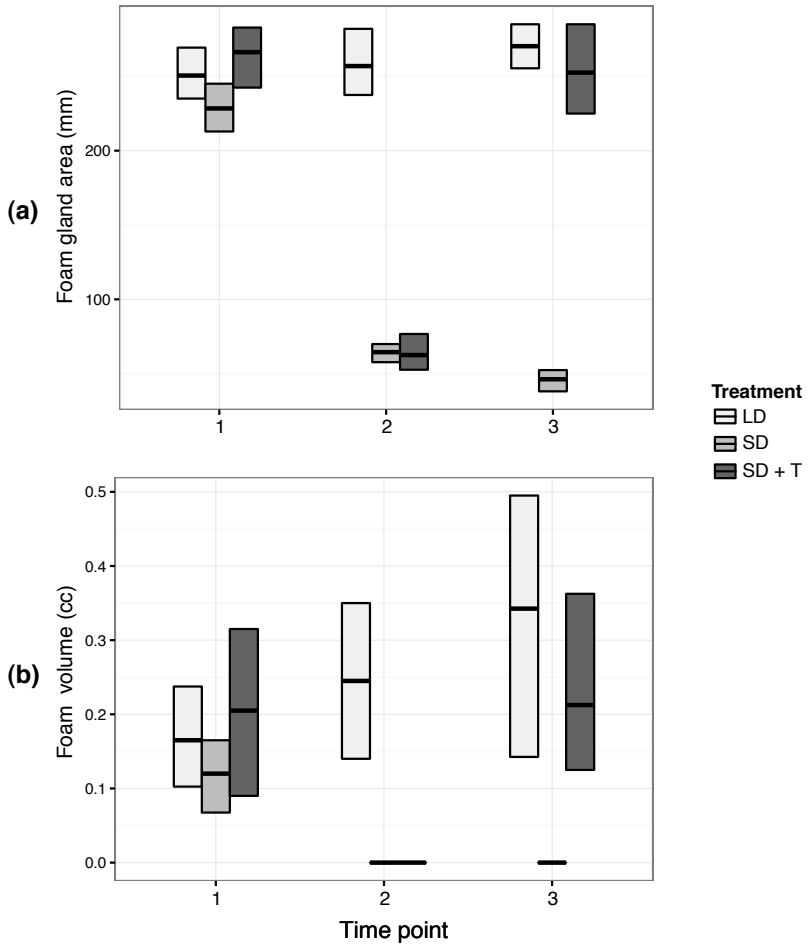

Figure 4.2 Foam glands responded to photoperiod and hormone treatments as expected. Average (bars) and 95% confidence intervals (boxes) of foam gland area (a) and volume (b) for Japanese quail males at three different time points. *Time points*: 1 = Baseline, 2 = After photoperiod treatment (*i.e.*, long or short days), 3 = After hormone administration. *Treatments*: LD = long day, SD = short day, SD + T = short day + testosterone. N equals 6 (a) and 4 (b) per treatment as volume measurements were only taken for Group 2 (see *Methods)*. CIs derived from bootstrap resamplings of the mean without assuming normality.

regulated when foam is produced *and* present in the proteomic dataset eliminated many genes with high protein abundances and expression levels. Therefore, we retained both groups for downstream analyses. We annotated them jointly, but analyzed them separately for the evolutionary analyses.

We report the 20 most abundant foam proteins (*P + GE*, *P*) according to emPAI in Table 4.1. We clustered annotations of biological processes and molecular functions according to term similarity. When compared to transcriptome expectations, genes encoding foam proteins (*P + GE*, *P*) were highly enriched for biological processes involving glycolysis and carbohydrate metabolism (Table 4.2: Annotation clusters 1 and 2). *P + GE* and *P* genes were also significantly enriched (*i.e.,* enrichments score > 1.3) for biological processes associated with the regulation of actin cytoskeleton organization and filaments, protein glycosylation of the N terminus, and protein localization (Table 4.2). Genes encoding foam proteins were significantly enriched for molecular functions encompassing GTP binding, peptidase inhibitor activity, nucleotide binding, and intramolecular oxidoreductase activity (Table 4.2).

High proportion of orthologs for foam proteins

Genes that encode foam proteins (*P + GE*, *P*) have relatively more orthologs in chicken than other classes of genes (Figure 4.4). This pattern was consistent if comparisons were made among genes categorized by their expression patterns in the foam gland (Figure 4.4a) or across different tissues (Figure 4.4b). Although genes that encoded foam proteins determined by both proteomic analysis and differential expression (*P+ GE*) had more orthologs in chicken than those identified only with proteomics (*P*), both groups

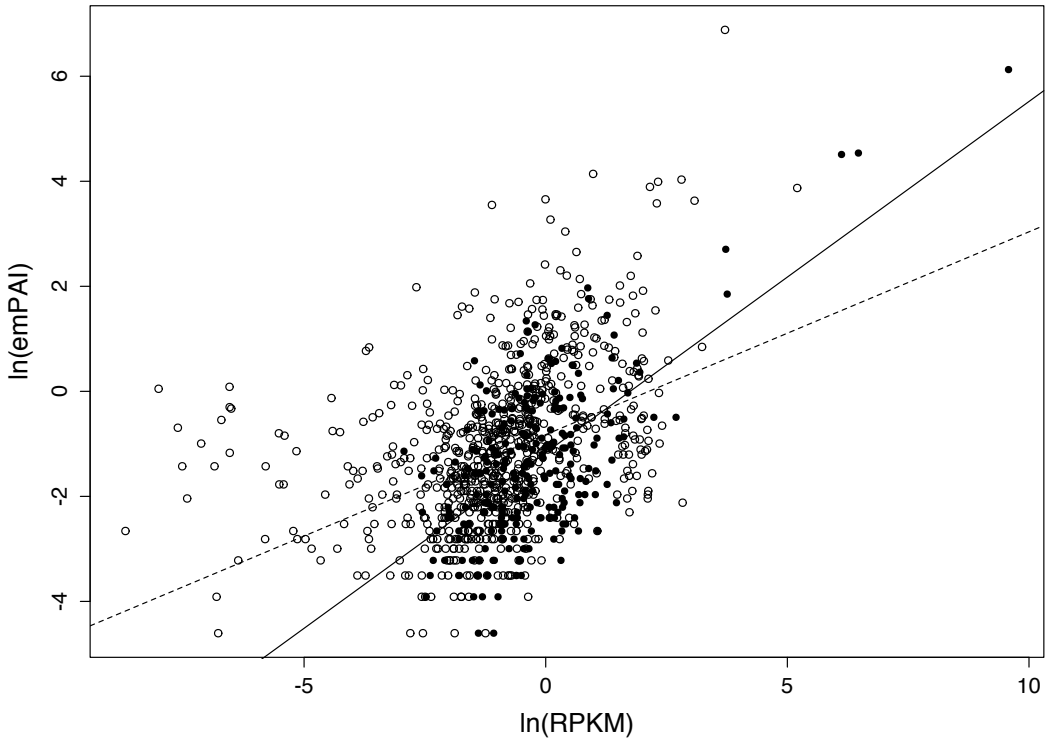

Figure 4.3 Protein abundance (emPAI) and expression level (RPKM) when the foam gland is producing foam are correlated. A significant relationship existed whether considering all genes encoding foam proteins identified proteomically ($r^2 = 0.1906$, $F_{1, 997} = 234.7$, $P < 2.2 * 10^{-16}$; dashed line) and the subset of those identified by both differential gene expression and proteomics ($r^2 = 0.4209$, $F_{1,251} = 169.4$, $P < 2.2 * 10^{-16}$; solid line). Open circles = 999 foam genes identified proteomically (*P*), closed circles = 253 foam genes identified by both proteomic and differential gene expression analyses (*P + GE*). Those genes with RPKM = 0 (N = 3) were removed from the dataset prior to analysis.

had much larger values than all other gene classes. Genes that were restricted to one tissue (*FG*, *Testis*, or *Liver*) had the fewest orthologs. *P* - values for all pairwise Fisher's exact tests can be found in .4 and S4.5.

Foam proteins evolve slowly

Genes that encode foam proteins (*P + GE*, *P*) had slower rates of protein evolution (ω) than most other classes of genes (non-overlapping CIs; Figure 4.5). *P + GE* and *P* genes displayed similar average ω values. *P + GE* and *P* genes evolved more slowly than genes expressed in the foam gland (*FG Exp*) and genes not expressed in the foam gland (*Other*), but at the same rate as genes that are up-regulated when the foam gland produces foam but do not encode proteins found in the foam (*GE*; Figure 4.5a). When comparing across different tissues, genes encoding foam proteins also exhibited very low rates of protein evolution (Figure 4.5b). Estimates of ω were significantly lower for *P + GE* and *P* genes than testis- or liver-restricted genes, but no different from genes specific to the foam gland that do not encode foam proteins (*FG*). *P* genes evolved slightly, though significantly, slower than ubiquitously expressed genes, with *P + GE* displaying a non-significant trend in the same direction.

135 foam proteins had a predicted signal peptide, and 84 of these also had orthologs with an ortholog hit ratio > 0.9. Although foam proteins evolve slowly relative to other protein classes, the subset of foam proteins that are secreted evolve relatively faster than those that are not secreted (Figure 4.6). The means and CIs indicate that secreted foam proteins evolve at rates similar to liver- and foam gland-specific proteins (Figure 4.5). The proportion of genes with significant shifts in ω along the quail lineage (either accelerations

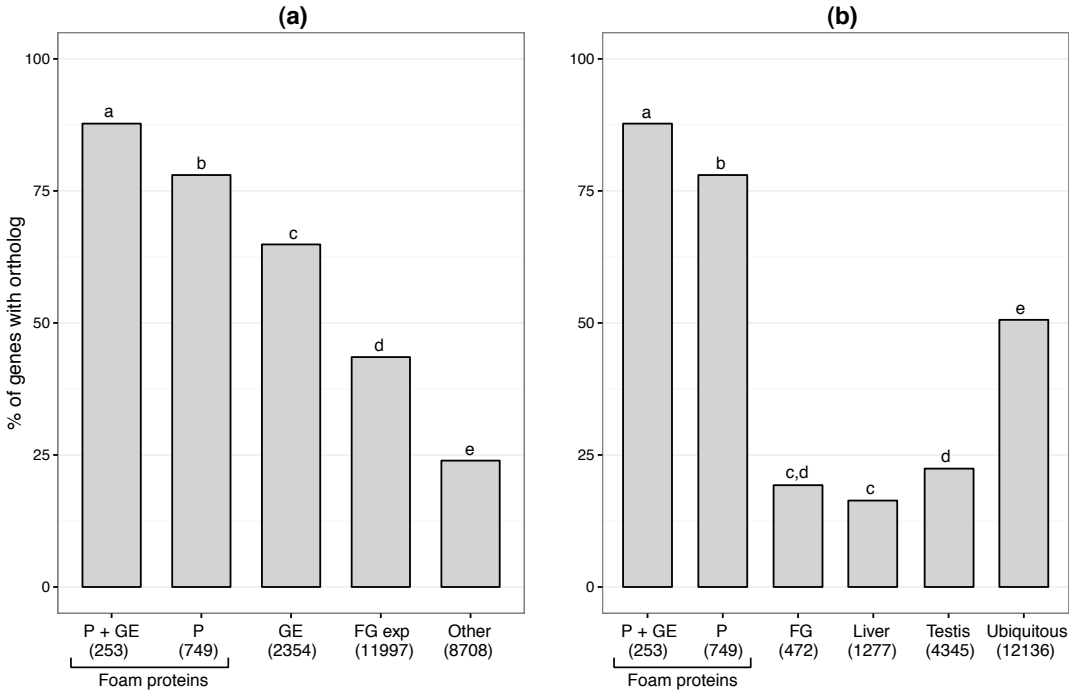

Figure 4.4. The foam proteome (*P + GE*, *P*) is disproportionately represented by orthologs. The percent of genes **(a)** expressed differentially in the foam gland or **(b)** from different tissues with 1:1 orthologs in chicken. Bars with unique letters are significantly different based on pairwise Fisher's exact tests ($P < 0.05$ after correcting for multiple comparisons). Samples sizes are given in parentheses and *P* - values for all tests are given in Table S4.4 & S4.5. *P + GE* = genes encoding foam proteins as determined by both proteomics and differential gene expression analyses; *P* = genes encoding foam proteins determined only proteomically; *GE* = genes up-regulated when the foam gland is producing foam; *FG Exp* = genes expressed in the foam, but not in the preceding categories; *Other* = genes not expressed in the foam gland; *FG* = Foam gland. In both (a) and (b), each gene was represented one time in the least inclusive appropriate category.

Table 4.1. Twenty most abundant foam proteins, annotations, and evolutionary rates.

Transcript	*G. gallus* ortholog[a,b]	Protein description[b,c]	Panther protein class[b,d]	emPAI[e]	RPKM[f]	Log FC foam[h]	ID method[i]	Signal peptide?[j]	dN[k]	dS[k]	dN/dS[k]	Ortholog hit ratio[l]	Alignment length (nuc)	Tissue-restricted[m]	Tissue-enriched[m]
Coja12182_c0_seq1	ENSGALG00000016761	Lysozyme, g-type 2		975.2	40.84	0.194	P	Y	0.074	0.155	0.475	0.995	633	U	FG
Coja17575_c0_seq1	No hit	Novel		457.9	14420.80	**3.000**	P + GE	Y						U	FG
Coja10321_c0_seq1	*ENSGALG00000023422*	*Uncharacterized protein*		93.44	645.61	**3.937**	P + GE	N				1.041		U	FG
Coja15808_c0_seq1	ENSGALG00000016962	Uncharacterized protein		90.95	455.29	**3.770**	P + GE	Y				0.849		U	FG
Coja10892_c0_seq2	ENSGALG00000013072	Aldo-keto reductase family 1, member B1	reductase	62.83	2.67	1.059	P	N	0.022	0.108	0.200	0.997	951	U	FG
Coja4845_c0_seq1	ENSGALG00000006453	Ovotransferrin (conalbumin)	transfer/carrier protein; serine protease; serine protease	56.36	16.60	-0.284	P	Y	0.109	0.226	0.482	1.001	2106	U	M
Coja4828_c0_seq1	ENSGALG00000001381	Actin, cytoplasmic type 5	actin and actin related protein	54.06	10.23	-0.243	P	N				0.997		U	
Coja21541_c0_seq1	ENSGALG00000003770	Annexin A2	transfer/carrier protein; annexin	49.06	8.64	-0.285	P	N	0.005	0.074	0.062	0.997	1017	U	FG
Coja12258_c0_seq1	*ENSGALG00000003506*	*Similar to ovoinhibitor precursor*	*enzyme modulator*	48.04	182.02	-0.918	P	N	0.082	0.247	0.331	1.000	1413	U	M
Coja10353_c0_seq1	Not ortholog			38.71	0.99	-0.040	P	N						FG	FG
Coja18772_c0_seq1	ENSGALG00000024011	Uncharacterized protein	transfer/carrier protein; isomerase	37.72	21.74	-0.039	P	Y				0.872		U	FG
Coja22380_c0_seq1	ENSGALG00000015148	Annexin A1	transfer/carrier protein; annexin	35.86	9.96	0.459	P	N	0.018	0.097	0.184	0.997	1026	U	FG
Coja7727_c0_seq1	Not ortholog			34.79	0.33	0.181	P	N						U	
Coja17586_c0_seq1	ENSGALG00000020180	Albumin (serum type)	transfer/carrier protein	26.32	1.10	-0.304	P	Y	0.062	0.140	0.438	0.997	1845	U	L
Coja8448_c0_seq1	ENSGALG00000000620	Fatty acid binding protein 3	transfer/carrier protein	20.95	1.50	-0.592	P	N	0.012	0.114	0.103	0.991	327	U	FG
Coja10336_c0_seq1	ENSGALG00000000919	Polymeric immunoglobulin receptor	immunoglobulin receptor superfamily; immunoglobulin receptor superfamily	14.95	41.49	**1.397**	P + GE	Y	0.129	0.322	0.401	0.822	1908	U	M
Coja9345_c0_seq1	No hit	Novel		14.2	1.89	**-3.683**	P	N						U	FG
Coja4850_c0_seq1	ENSGALG00000014442	Glyceraldehyde-3-phosphate dehydrogenase	dehydrogenase	13.19	6.67	**-2.436**	P	N	0.006	0.155	0.038	0.997	975	U	
Coja12359_c0_seq1	ENSGALG00000003053	Peroxiredoxin 6	peroxidase	11.19	0.99	-0.155	P	N	0.005	0.166	0.033	0.996	672	U	

[a]Ortholog assigned by reciprocal best BLASTp between quail and chicken protein; annotations are based on chicken ortholog

[b]Entries in italics are based on Ensembl v. 69 annotations that were retired or changed in Ensembl v. 71.

[c]Protein descriptions are based on Ensembl v. 69 gene IDs or Wiki gene IDs

[d]Protein classes based on PANTHER ontology terms from Ensembl ID; Novel = no hit in BLAST; Not ortholog = BLASTp hit, but not identified as they ortholog

[e]Percent of protein covered by peptides identified proteomically

[e]Protein abundance estimated by the exponentially modified protein abundance index (Ishihama, 2005)

[f]Gene expression level estimated as reads per kilobased per million mapped reads

[h]Log fold changes reported from averages of SD vs SD+ T and SD vs LD treatments; Values in bold are signficantly up- or down- regulated

[i]P = proteomics, GE = gene expression analyses

[j]Presence of a signal peptide assigned in SignalP 4.1

[k]Pairwise estimates based on quail:chicken orthologs

[l]Predicted peptide length of a quail gene divided by the length of the chicken ortholog

[m]Enriched: significantly enriched in one tissue by at least log 2 fold change; Restricted = expressed only in one tissue or ubiquitous; FG = foam gland, L= liver, T = testis, U = ubiquitous, M = multiple tissues

Table 4.2: Enrichment[a] of clustered Gene Ontology (GO) biological processes and molecular functions in genes encoding foam proteins (*P + GE*,*P*)

GO Term	Description	Enrichment Score	# genes[b]	%[c]	*P*-value[d]	Benjamini[e]	Term class[f]
Annotation cluster 1: monosaccharide metabolic process		7.31					
GO:0005996	monosaccharide metabolic process		26	3.36	1.66E-08	7.67E-06	BP
GO:0019318	hexose metabolic process		24	3.10	4.19E-08	1.55E-05	BP
GO:0006006	glucose metabolic process		20	2.59	1.62E-07	5.01E-05	BP
Annotation cluster 2: glycolysis and carbohydrate catabolic process		6.37					
GO:0016052	carbohydrate catabolic process		20	2.59	4.00E-09	2.47E-06	BP
GO:0006006	glucose metabolic process		20	2.59	1.62E-07	5.01E-05	BP
GO:0044275	cellular carbohydrate catabolic process		16	2.07	8.05E-07	2.13E-04	BP
GO:0046164	alcohol catabolic process		16	2.07	8.05E-07	2.13E-04	BP
GO:0006096	glycolysis		13	1.68	1.06E-06	2.46E-04	BP
GO:0019320	hexose catabolic process		15	1.94	1.15E-06	2.12E-04	BP
GO:0006007	glucose catabolic process		15	1.94	1.15E-06	2.12E-04	BP
GO:0046365	monosaccharide catabolic process		15	1.94	1.97E-06	3.32E-04	BP
Annotation cluster 3: GTP binding		5.00					
GO:0032561	guanyl ribonucleotide binding		35	4.53	9.96E-06	9.76E-04	MF
GO:0019001	guanyl nucleotide binding		35	4.53	9.96E-06	9.76E-04	MF
GO:0005525	GTP binding		35	4.53	9.96E-06	9.76E-04	MF
Annotation cluster 4: regulation of cytoskeleton organization		3.01					
GO:0051493	regulation of cytoskeleton organization		13	1.68	7.41E-05	0.833	BP
GO:0032956	regulation of actin cytoskeleton organization		10	1.29	1.03E-04	0.014	BP
GO:0030832	regulation of actin filament length		9	1.16	2.56E-04	0.028	BP
GO:0008064	regulation of actin polymerization or depolymerization		9	1.16	2.56E-04	0.028	BP
GO:0032970	regulation of actin filament-based process		10	1.29	2.84E-04	0.029	BP
GO:0030833	regulation of actin filament polymerization		8	1.03	6.40E-04	0.052	BP
GO:0032271	regulation of protein polymerization		8	1.03	0.003	0.158	BP
GO:0044087	regulation of cellular component biogenesis		10	1.29	0.005	0.240	BP
GO:0043254	regulation of protein complex assembly		8	1.03	0.010	0.414	BP
GO:0032535	regulation of cellular component size		11	1.42	0.063	0.800	BP
Annotation cluster 5: actin cytoskeleton organization		2.67					
GO:0030036	actin cytoskeleton organization		13	1.68	3.19E-04	0.031	BP
GO:0030029	actin filament-based process		13	1.68	5.95E-04	0.051	BP
GO:0007010	cytoskeleton organization		15	1.94	0.052	0.804	BP
Annotation cluster 6: regulation of actin filament		2.06					
GO:0030834	regulation of actin filament depolymerization		6	0.78	0.001	0.080	BP
GO:0030835	negative regulation of actin filament depolymerization		5	0.65	0.003	0.157	BP
GO:0030837	negative regulation of actin filament polymerization		5	0.65	0.003	0.157	BP
GO:0051693	actin filament capping		5	0.65	0.003	0.157	BP
GO:0043244	regulation of protein complex disassembly		6	0.78	0.010	0.416	BP
GO:0032272	negative regulation of protein polymerization		5	0.65	0.011	0.411	BP
GO:0031333	negative regulation of protein complex assembly		5	0.65	0.017	0.516	BP
GO:0043242	negative regulation of protein complex disassembly		5	0.65	0.017	0.516	BP
GO:0051494	negative regulation of cytoskeleton organization		5	0.65	0.037	0.736	BP
GO:0010639	negative regulation of organelle organization		5	0.65	0.083	0.892	BP
Annotation cluster 7: peptidase inhibitor activity		2.01					
GO:0004857	enzyme inhibitor activity		14	1.81	0.007	0.227	MF
GO:0004866	endopeptidase inhibitor activity		11	1.42	0.010	0.275	MF
GO:0030414	peptidase inhibitor activity		11	1.42	0.012	0.297	MF
Annotation cluster 8: nucleotide binding		1.76					
GO:0017076	purine nucleotide binding		112	14.49	0.011	0.484	MF
GO:0032555	purine ribonucleotide binding		106	13.71	0.014	0.761	MF
GO:0032553	ribonucleotide binding		106	13.71	0.014	0.761	MF
GO:0000166	nucleotide binding		126	16.30	0.042	0.761	MF
Annotation cluster 9: protein localization		1.47					
GO:0045184	establishment of protein localization		32	4.14	0.027	0.649	BP
GO:0015031	protein transport		32	4.14	0.027	0.649	BP
GO:0008104	protein localization		34	4.40	0.052	0.808	BP
Annotation cluster 10: protein amino acid N-linked glycosylation		1.53					
GO:0006487	protein amino acid N-linked glycosylation		4	0.52	0.015	0.484	BP
GO:0018196	peptidyl-asparagine modification		3	0.39	0.042	0.761	BP
GO:0018279	protein amino acid N-linked glycosylation via asparagine		3	0.39	0.042	0.761	BP
GO:0004576	oligosaccharyl transferase activity		3	0.39	0.071	0.718	MF
Annotation cluster 11: intramolecular oxidoreductase activity		1.42					
GO:0016862	intramolecular oxidoreductase activity, interconverting keto- and enol-groups		3	0.39	0.038	0.559	MF
GO:0016864	intramolecular oxidoreductase activity, transposing S-S bonds		3	0.39	0.038	0.559	MF
GO:0003756	protein disulfide isomerase activity		3	0.39	0.038	0.559	MF

[a]Term clustering and enrichment analysis performed in DAVID functional annotation tool v. 6.7

[b]Based on the number of *G. gallus* orthologous genes with a particular GO term; enrichment scores > 1.3 correspond to a *P*- value < 0.05

[c]Percent of *G. gallus* orthologs of genes encoding foam proteins ('P + GE' and 'P') that have a particular GO term

[d]From a modified Fisher's exact test (EASE score) testing for overrepresentation of a GO term in the foam gene list versus the transcriptome

[e]Global correction for *P*- values using the Benjamini multiple testing correction as implemented in DAVID

[f]BP = biological process, MF = molecular function

or decelerations) did not differ across foam gland or tissue class *(*Fisher's exact test: $P > 0.05$ in all comparisons).

Foam proteins have more orthologs in phylogenetically distant species

Overall, genes that encode foam proteins were more likely than other genes to have orthologs with evolutionarily distant groups (Figure 4.7). Both sets of candidate genes encoding foam proteins (*P + GE*, *P*) displayed a higher proportion of genes with origins in bacteria and/or archaea than genes just up-regulated in the active foam gland (*GE*), expressed in the foam gland (*FG Exp*) or not expressed in the foam gland (*Other*). In both analyses, *P* and *P + GE* genes showed generally the same trends. *P* - values for all pairwise Fisher's exact tests can be found in Tables S4.6 and S4.7.

Discussion

We present a comprehensive analysis of the molecular underpinnings of a novel avian copulatory fluid. Using standard proteomic techniques, we identified 1002 genes that encode proteins present in the unique copulatory foam of male Japanese quail (Tables 4.1, S4.3). Because foam may include cellular debris and other contaminants, we also compared gene expression profiles of foam glands in their sexually active and inactive states. By combining the gene expression and proteomics datasets, we identified 253 proteins that we have confidence are "true" structural/functional components of foam. Our integrative approach is straightforward and applicable across many systems, as many organisms seasonally regress their reproductive tissues.

Potential role of foam proteins

In the section below, we explore the potential role of foam proteins based on well-annotated *G. gallus* orthologs and the functions of these proteins in the seminal fluid of other species. Our differential gene expression analysis was conservative and missed several highly abundant foam proteins (Figure 4.2). This is likely because foam proteins expressed at high levels in a regressed foam gland may not meet the threshold for differential expression. Therefore, we interpret proteins that meet one of the following criteria as foam proteins: 1) those identified both proteomically and by differential gene expression analyses (*P + GE*) and 2) proteins with a predicted signal peptide and high abundance at both the protein and gene expression level (Table 4.1). We limit our discussion of the potential biological function to these groups of foam proteins. We also discuss biological processes and molecular functions that are enriched in foam (Table 4.2).

Antibiotic properties

Bacteria in male or female reproductive tracts can directly damage sperm cells, resulting in reduced sperm viability and motility (Moretti *et al.*, 2008; Kaur *et al.*, 2010; Otti *et al.*, 2012). Ejaculates commonly possess antibiotic substances transferred to females during mating, presumably to limit the detrimental effects of sperm-associated bacteria (Poiani, 2006; Avila *et al.*, 2011). One such substance, lysozyme, contributes to antibacterial defense by cleaving key structural components of bacterial cell walls. When present in seminal fluid, lysozyme can even reduce sperm mortality caused by microbes (Otti *et al.*, 2012). Goose-type lysozyme 2 (hereafter 'lysozyme *g*') is by far the most abundant protein in foam (Tables 4.1 & S4.3). Although it is not significantly up-regulated

when foam is being produced, it is highly expressed in active foam glands and contains a signal peptide. Thus, we are confident that lysozyme *g* is a major constituent of foam.

While lysozyme or lysozyme-like activity has been described in SFPs before, its presence is distinct in foam in two ways (Poiani, 2006; Avila *et al.*, 2011). First, lysozyme is the dominant foam protein. When described in the ejaculates of other species, lysozyme is one of many components of intermediate abundance (*e.g.*, S*. Dorus*, *unpublished data*; Bourgeon, 2003). The dual excretory and reproductive nature of the avian cloaca could permit gut microbes to be incorporated into ejaculates providing more opportunity for pathogen exposure and increased selection for antimicrobial defenses (Sheldon, 1993). As such, substances with antibiotic properties may be strongly favored in the ejaculates of avian taxa. Indeed, variation in lysozyme bactericidal activity exists in the seminal fluid of wild birds and antibacterial activity of semen is likely a target of natural or sexual selection (Rowe *et al.*, 2011; 2013).

Second, lysozyme *c,* not lysozyme *g*, is present in chicken seminal fluid (*S. Dorus*, unpublished data). Although both types of lysozyme display similar activity and three-dimensional structures, they have radically different amino acid sequences (Callewaert & Michiels, 2010). Interestingly, lysozyme *g* is a key component in the egg white of goose, ducks, and other birds, whereas lysozyme *c* is dominant in the albumen of chicken and other Galliformes (presumably Japanese quail; Prager *et al.*, 1974; Callewaert & Michiels, 2010). Although lysozyme *g* is expressed in other tissues in chicken (*e.g.*, bone marrow, lung), it is not expressed in the chicken oviduct where egg white is produced (Nakano & Graf, 1991). Both egg white albumen and the foam gland secretion can be whipped into a meringue-like substance. Thus, lysozyme *g* may be an interesting example of co-option,

whereby *Coturnix* males re-purposed a major egg white protein utilized by other taxa to contribute similar properties in the foam.

In addition to lysozyme *g,* several other substances in the foam display antibiotic activity. These include ovotransferrin and protease inhibitors such as cystatin, ovoinhibitor, and ovalbumin (Ibrahim *et al.*, 2000; Bourgeon, 2003; Pellegrini *et al.*, 2004; Wesierska *et al.*, 2005; Bourin *et al.*, 2011). Thus, one function of foam appears to be to rid the female reproductive tract of pathogens that can harm sperm.

Facilitation of aerobic respiration

We clustered biological processes according to similarity to determine which processes are overrepresented in foam proteins compared with those from the entire transcriptome. The two most highly enriched annotation clusters involved carbohydrate metabolism: monosaccharide metabolic and glycolysis/carbohydrate catabolic processes (Table 4.2). The enriched clusters point to a pivotal role that foam could play in meeting the metabolic demands of quail spermatozoa.

Maintaining sperm motility is energy-demanding, yet crucial for male reproductive success. Energy for sperm motility is generally derived via one of two metabolic pathways: oxidative phosphorylation or glycolysis (Turner, 2003). These two processes are generally partitioned across the sperm. Oxidative phosphorylation occurs in the mitochondria located in sperm midpieces, while glycolysis takes place primarily in enzymes bound to the sperm flagellum (Turner, 2003; Ford, 2006). The results of our enrichment analysis suggest that foam might provide an additional site for glycolysis and carbohydrate metabolism. In the presence of oxygen, pyruvate, the end product of glycolysis, enters the

citric acid cycle in the mitochondrial matrix, which in turn fuels oxidative phosphorylation, thereby generating ATP and energy (Visconti, 2012). Our data suggest a model whereby foam facilitates aerobic respiration of sperm by providing 1) enzymes that metabolize carbohydrates, byproducts (*e.g.*, pyruvate) of which can be utilized by the mitochondria in the sperm midpiece as substrates for the citric acid cycle and 2) an oxygenated environment.

We emphasize that our 'foam facilitates sperm aerobic respiration' model is speculative at this point and requires further investigation. However, several lines of evidence are consistent with our hypothesis. Pyruvate, the end product of glycolysis, and lactate alone can maintain basic metabolic functions of Japanese quail sperm (Singh *et al.*, 2011b). Lactic dehydrogenase is the enzyme that catalyzes the interconversion of lactate and pyruvate. The seminal fluid of Japanese quail exhibits ten times the lactic dehydrogenase activity than that of chicken or turkey and may utilize lactate, possibly also supplied by foam, as a substrate (Buxton & Orcutt, 1975; Singh *et al.*, 2011a; Singh *et al.*, 2011). Moreover, simple aeration improves the fertility of stored turkey and chicken semen (Wishart, 1981). If foam enables sperm to preferentially utilize aerobic respiration, it could provide a link between observed improvements in sperm motility and viability and the presence of foam (Cheng *et al.*, 1989b; Singh *et al.*, 2011a). The proposed function of foam could provide fitness benefits, as aerobic respiration is more efficient and produces more ATP than anaerobic respiration. In Galliformes, sperm ATP content is correlated with sperm mobility and the fertilizing ability of sperm (Wishart, 1982; Froman & Feltmann, 1998).

If foam confers selective advantages through improved aerobic respiration of sperm, why is it restricted to *Coturnix* quail? Unique features of Japanese quail sperm suggest they may have unusually high metabolic demands. Among Galliformes, Japanese quail sperm possess a flagellum over twice as long as that found in other non-passerine birds (Korn *et al.*, 2000). Quail sperm also have exceptionally long midpieces covering 64-74% of total sperm flagellum length (Woolley, 1995; Korn *et al.*, 2000). These midpieces are packed with between 1,400 - 2,500 mitochondria per sperm (Woolley, 1995; Korn *et al.*, 2000). Interestingly, sperm axoneme proteins as a group evolve adaptively in quail and experienced a shift towards accelerated rates of evolution sometime after quail and chicken split (Finseth *et al.*, 2013a). Foam, therefore, could have provided a biochemical environment that promoted initial diversification of sperm structure. Alternatively, the evolution of foam may have been a response to increased energetic demands imposed by novel aspects of Japanese quail sperm.

Suppression of female immune response

For sperm, the female reproductive tract can be a remarkably hostile environment (Birkhead *et al.*, 1993). In fowl, less than 1% of inseminated sperm make it to the sperm storage tubules, the site of long term storage in the female (Bakst, 2011). Heavy sperm selection takes place in the avian vagina and, as sperm can be recognized as foreign by the female immune system, selection may be directly mediated by a female immune response (Birkhead *et al.*, 1993; Poiani, 2002; Das *et al.*, 2008). In Japanese quail, the number of leukocytes and lymphocytes increases in the vagina soon after mating, and many of these immune cells are closely associated with spermatozoa (Higaki *et al.*, 1995; Das *et al.*, 2008).

Males that provide their sperm with substances that permit the evasion of the female immune system are likely to possess a fertilization advantage.

The role of SFPs in immunosuppression of the female reproductive tract is best described in mammals, and several proteins described in these studies are also found in foam (James & Hargreave, 1984; Robertson, 2007; Robertson *et al.*, 2009). Chief among these is ovotransferrin, a serine protease that binds iron and is a major constituent of foam. The human ortholog of ovotransferrin, lactotransferrin, reduces the female immune response by inhibiting lymphocyte proliferation in the reproductive tract (reviewed in Poiani, 2006). In fact, lactotransferrin plays a dual role in sperm defense by suppressing the female immune system and providing antimicrobial properties (Weinberg, 2001; Poiani, 2006). Some other major and minor foam proteins that are mammalian SFPs capable of inhibiting the female immune response include proteases (*e.g.*, HtrA serine peptidase 1), cytokines (*e.g.*, Interleukin 1 receptor accessory protein, Interleukin 4 induced 1), immunoglobulins (Ig lambda chain C region), and transglutaminase (James & Hargreave, 1984; Poiani, 2006). Interestingly, transglutaminase gives mouse mating plugs their viscosity and may also 'hide' the antigenic properties of sperm from female immune cells (James & Hargreave, 1984; Dean, 2013). Consistent with the idea that foam mediates immunosuppression, we mated females to males with foam but not sperm and found that the females' vaginas increased levels of transforming growth factor β 1 receptor (*TGFBR1*, unpublished data). In mammals, *TGFBR1* is part of a pathway that suppresses the proliferation of T- and B- lymphocytes (Wahl *et al.*, 1988; Robertson *et al.*, 2002; Robertson, 2005; Robertson *et al.*, 2009). Foam may, therefore, play an important role in immunosuppression, suggesting possible costs in females.

One of the most enriched molecular function clusters in foam is peptidase inhibitor activity (Annotation cluster 7, Table 4.2). Ejaculated mouse proteins are also significantly enriched for peptidase inhibitors (Dean *et al.*, 2011). In mice, females up-regulate proteases in response to mating (Dean *et al.*, 2011). Peptidase inhibitors in ejaculates have been hypothesized to protect sperm from a female-mounted proteolytic attack or prevent degradation of the copulatory plug by female proteases (Dean *et al.*, 2009). Either way, peptidase inhibitors deriving from the seminal fluid protect male reproductive interests and foam may confer similar benefits. If female interests differ from males, peptidase inhibitors in foam could be targets of sexual conflict.

Conservation of foam proteins as a class

SFPs frequently exhibit rapid molecular evolution. In *Drosophila* alone, protein sequences of SFPs diverge rapidly and sometimes under positive selection (Civetta & Singh, 1995; Swanson *et al.*, 2001b; Wagstaff, 2005), individual SFPs experience rapid birth/death processes (Mueller *et al.*, 2005; Begun *et al.*, 2006b; Findlay *et al.*, 2008), duplication of SFPs is sometimes followed by adaptive divergence (Mueller *et al.*, 2005; Wagstaff & Begun, 2007; Findlay *et al.*, 2008; Almeida & DeSalle, 2008), and new SFPs arise from unknown evolutionary origins (Begun *et al.*, 2006a; Findlay *et al.*, 2009). These dynamic molecular evolutionary processes produce at least two major patterns: a high incidence of novel genes (SFPs without detectable orthologs; *e.g.*, Kelleher *et al.*, 2009; Walters & Harrison, 2010) and elevated estimates of ω in comparison to other types of genes (*e.g.*, Walters & Harrison, 2010; Dorus *et al.*, 2006; Dean *et al.* 2009).

Contrary to the expectation of rapid evolution, we observe a striking degree of conservation among putative foam proteins as a class. Four lines of evidence allow us to draw this conclusion: *i*) foam proteins reveal significantly lower rates of protein evolution (ω) than widespread genes or genes expressed in other tissues (Figure 4.5), *ii*) foam proteins have more orthologs than other gene classes (Figure 4.4), *iii*) these orthologs disproportionately derive from phylogenetically distant groups (Figure 4.7) and *iv*) we see no evidence of major shifts in selective pressure along the *Coturnix* lineage (see Results: *Foam proteins evolve slowly*). Molecular evolutionary patterns are quite similar between proteins identified only proteomically (*P*) and those identified using both proteomics and change in gene expression (*P + GE, P*). These findings suggest that the novel foam proteome arose in large part by co-option of ancestral genes followed by sequence conservation (True & Carroll, 2002). Our results are consistent with previous work showing that the genes encoding the foam gland also evolve slowly (Finseth *et al.* 2013a).

This raises the question why foam appears to rely so heavily on re-purposing of conserved proteins, when SFPs are generally evolutionary labile. The foam proteome differs from that of typical seminal fluids in several fundamental ways. Foam is a novel proteome that is restricted to *Coturnix* quail (Klemm *et al.*, 1973). Co-option of ancestral genes is a major theme of the evolution of novel phenotypes, with divergence in regulatory elements sometimes being more important than other genetic changes (True & Carroll, 2002; Wray, 2007; Carroll, 2008). For example, genes with ancient evolutionary origins

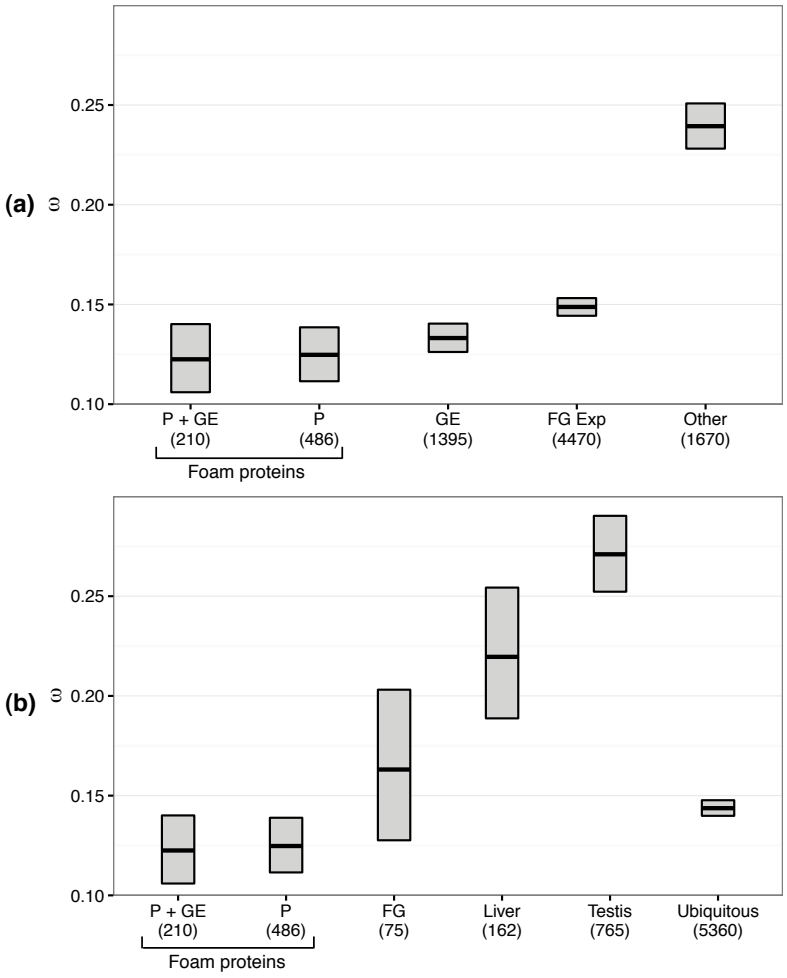

Figure 4.5 Foam proteins (*P + GE*, *P*) evolve slowly. Mean pairwise ω estimates (bar) and 95% CIs (boxes) calculated from 1:1 orthologs between quail and chicken. Ortholog pairs were classified either **(a)** based on their relative expression in the foam gland or **(b)** as expressed in only one of three tissues or ubiquitously expressed. CIs derived from bootstrap resamplings of the mean without assuming normality. Samples sizes are given in parentheses. *P + GE* = genes that encode foam proteins as determined by both proteomics and differential gene expression analyses; *P* = genes that encode foam proteins determined only proteomically; *GE* = genes up-regulated when the foam gland is producing foam; *FG Exp* = genes expressed in the foam, but not in the preceding categories; *Other* = genes not expressed in the foam gland; *FG* = Foam gland. Expression determined in Finseth *et al*. 2013a. In both (a) and (b), each gene was represented one time in the least inclusive appropriate category.

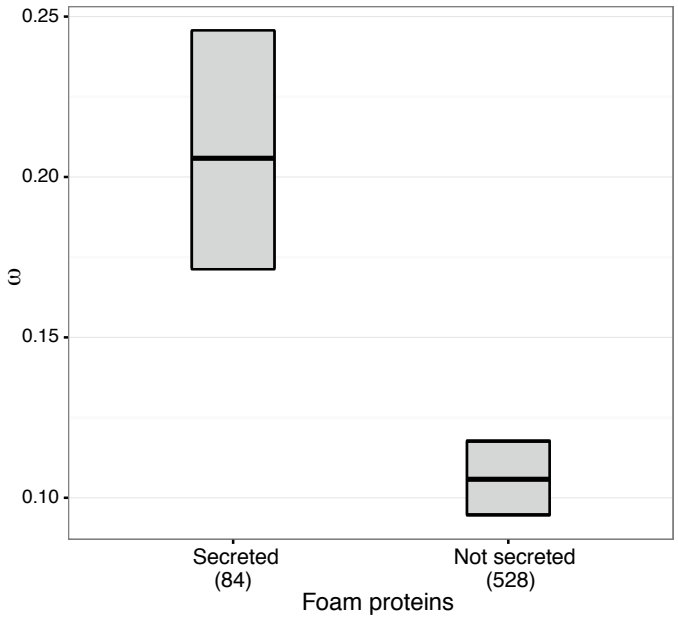

Figure 4.6 Secreted foam proteins evolve relatively faster than those with no evidence for secretion. Mean pairwise ω estimates (bar) and 95% CIs (boxes) calculated from 1:1 orthologs between quail and chicken. Secretion was determined by the presence of a signal peptide. CIs derived from bootstrap resamplings of the mean without assuming normality. The set of foam proteins analyzed were those from the entire proteomics (*P*) dataset. Samples sizes are in parentheses.

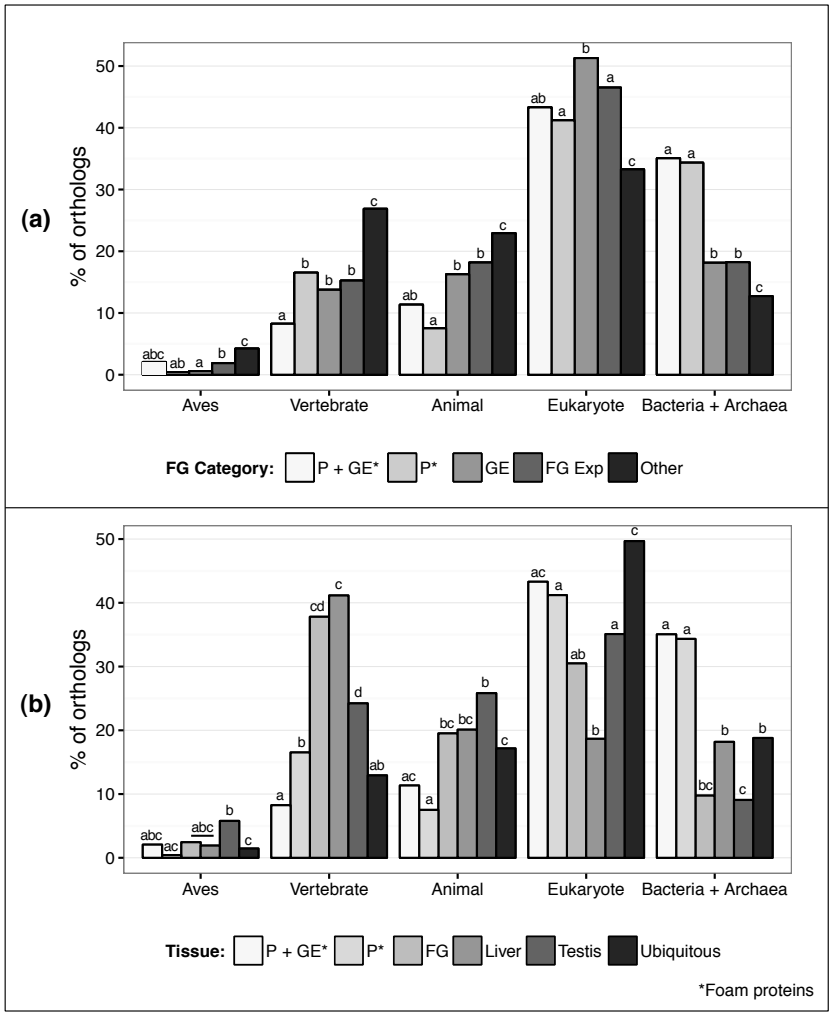

Figure 4.7 Foam proteins are disproportionately represented by genes with orthologs in distant ancestors. Evolutionary origins of genes according to **(a)** relative expression in the foam gland or **(b)** across different tissues. The most distantly related species with an ortholog served as a proxy for the evolutionary origin of a gene. Analyses were restricted to those genes with 1:1 orthologs in the chicken. For (b), genes categorized as FG (N = 82), liver (N = 209), or testis (N = 1003) were only expressed in those tissues, while ubiquitous genes were expressed in all three tissues (N = 5438). Bars with unique letters are significantly different based on pairwise Fisher's exact tests performed within each 'origin' group. *P + GE* = genes encoding foam proteins as determined by both proteomics and differential gene expression analyses (N = 194); *P* = genes encoding foam proteins determined only proteomically (N = 466); *GE* = genes up-regulated when the foam gland is producing foam (N = 1,396); *FG Exp* = genes expressed in the foam gland (N = 4,652); *Other* = genes not expressed in the foam gland (N = 2129); *FG* = Foam gland. *P* - values for all tests are given in Tables S4.6 & S4.7.

were co-opted during the development of the placenta, likely due to modifications at regulatory elements (Knox & Baker, 2008). Increased pleiotropic constraints on foam proteins due to novel spatiotemporal expression patterns in the foam gland could have intensified purifying selection of foam proteins, further reducing rates of evolution. While we do not observe an excess of decelerations of foam proteins at the class level (see Results: *Foam proteins evolve slowly*), increased pleiotriopic constraints may have been an important selective force on individual genes.

It is possible that the aerated structure of foam, rather than particular chemical constituents, confers fertility benefits to males. This would be in contrast with models of SFP evolution, which generally implicate interactions with specific molecules in antagonistic or cooperative ways to facilitate rapid evolution (Wolfner & Wolfner, 2009). Japanese quail sperm clump easily and foam disaggregates sperm upon contact, which could be due simply to the structural matrix formed by foam (Singh *et al.*, 2011a). The bubbles in foam may also provide an oxygenated environment to sperm, thereby improving aerobic respiration (see Discussion: *Facilitation of aerobic respiration*). Many major foam proteins are also in the egg white of chicken (*e.g.*, ovotransferrin, ovoinhibitor, cystatin, ovalbumin) or goose (*e.g.*, lysozyme, g-type) albumen (Prager *et al.*, 1994; Mann 2007). Egg white is also capable of producing a bubbly matrix from a similar whipping action (*i.e.* meringue). If foam functions mainly through emergent properties arising from new combinations/interactions of conserved proteins, diversifying selection on individual proteins may be weak.

Many hypotheses proposed to explain the rapid divergence of reproductive proteins invoke antagonistic coevolutionary dynamics. While foam mediates sperm competition, it

does so through conferring fertilization benefits to sperm regardless of its source (Finseth *et al*. 2013b, Cheng *et al.*, 1989b). In other words, foam is only known to influence intrasexual competition through non-antagonistic interactions. Thus, a lack of conflict in sperm competition may underlie the observed conservation of foam proteins. However, male-derived foam proteins may play a role in sexual antagonism with females (see Discussion: *Suppression of female immune response*). Foam is also physically separated from the seminal fluid, not packaged with sperm inside males, and not required for fertilization, and the foam gland is not limited to males but found in rudimentary form in females (discussed in Finseth *et al*., 2013a). Any of these unique aspects of the Japanese quail-foam system could potentially influence selection on foam proteins.

Rapid divergence of important foam proteins

Although foam proteins as a class are conserved, two observations suggest that proteins that are integral to the function of the foam may be evolving rapidly. Two foam proteins, *g*-type lysozyme 2 (see Discussion: *Antibiotic properties*) and a novel protein (*Coja17575_c0_seq1*), are approximately an order of magnitude more abundant than any other foam protein (Table 4.2: emPAI). *Coja17575_c0_seq1* is significantly enriched when the foam gland is reproductively active, contains a predicted signal peptide, and its RNA transcript is expressed at levels more than 20X greater than those of any other foam protein. For these reasons, we believe this novel protein represents a real gene that encodes a major foam protein. *Coja17575_c0_seq1* exhibits no similarity to other described nucleotide or protein sequences. This novel protein may be either evolving so rapidly that

detectable sequence homology is obscured or a newly evolved gene of unknown origins (Long *et al.*, 2003). Either of these scenarios suggests dynamic evolutionary processes.

Additionally, foam proteins with predicted signal peptides exhibit elevated rates of protein evolution relative to those without one (Figure 4.6). Signal peptide presence is one criterion previous studies used when attempting to identify SFPs bioinformatically from expressed sequence data (*e.g.*, Swanson *et al.*, 2001b; Andres *et al.*, 2006; Kelleher *et al.*, 2009; Walters & Harrison, 2010). The presence of a signal sequence suggests that a protein is secreted from the cell. Secreted proteins are available for transfer to the female during mating and are able to interact with either the male or female reproductive tracts or sperm. Reproductive proteins that interact with the environment in this way have been hypothesized to have the greatest potential for adaptation in response to sexual selection (Dorus *et al.*, 2006; Dean *et al.*, 2009; Dorus *et al.*, 2010). Consistent with this model, secreted foam proteins evolve much faster than foam proteins that we have no evidence contain signal peptides. Thus, elevated divergence could be produced by positive selection, possibly on interacting secreted foam proteins as they adapt to intra- or inter-sexual selection. Alternatively, the relatively high rates of evolution may be a product of reduced selective constraint on secreted proteins or increased negative selection on non-secreted proteins.

Conclusion

We combined standard proteomic techniques with a RNA-Seq approach to identify the proteins comprising a novel copulatory fluid, the foam of Japanese quail. Foam proteins reveal contrasting patterns of constraint and divergence. On the one hand, as a class, foam

proteins exhibit conservation at the sequence level and include a disproportionate number of genes with phylogenetically distant orthologs. On the other hand, highly abundant and secreted foam proteins are evolutionarily dynamic. This paradox may be partly explained by the potential functions of foam proteins. Foam may foster aerobic respiration of sperm, largely through emergent properties of combinations of co-opted proteins. Other foam proteins potentially contribute antimicrobial properties or suppress the female immune system, and may evolve under antagonistic coevolutionary dynamics. Thus, novel structures and functions may emerge both as a result of the packaging of conserved proteins in new combinations together with the rapid evolution of individual components.

Acknowledgements

The authors would like to thank G. Findlay, S. Dorus, and the Harrison lab for helpful discussions and advice; E. Bondra for performing the RT-qPCR; E. Adkins-Regan for quail samples and housing; M. Mouchka, R. Bukowski and Q. Sun for assistance with transcriptome assembly and RNA-Seq data analysis; S. Zhang and J. McCardle for support with proteomics; J. Mosher, and P. Schweitzer for aiding with cDNA library preparation and Illumina sequencing; S. Iacovelli and N. Baran for quail handling; and T. van Deusen, S. Martin, L. Vann, and P. Smith for quail care. This research was funded by an NSF DDIG DEB-1010757 to FRF and RGH, P.E.O. Scholar, Cornell Sigma Xi, Andrew W. Mellon, and Paul. F. Feeny awards to FRF.

References

Adkins-Regan, E. 1999. Foam produced by male *Coturnix* quail: What is its function? *Auk* **116**: 184–193.

Almeida, F. & DeSalle, R. 2008. Evidence of adaptive evolution of accessory gland proteins in closely related species of the Drosophila repleta group. *Molecular Biology and Evolution* **25**: 2043–2053.

Andersen, C.L., Jensen, J.L. & Ørntoft, T.F. 2004. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer research* **64**: 5245–5250. AACR.

Andres, J., Maroja, L., Bogdanowicz, S., Swanson, W. & Harrison, R. 2006. Molecular evolution of seminal proteins in field crickets. *Molecular Biology and Evolution* **23**: 1574–1584.

Andrés, J.A., Maroja, L.S. & Harrison, R.G. 2008. Searching for candidate speciation genes using a proteomic approach: seminal proteins in field crickets. *Proceedings Of The Royal Society B-Biological Sciences* **275**: 1975–1983.

Avila, F.W., Sirot, L.K., Laflamme, B.A., Rubinstein, C.D. & Wolfner, M.F. 2011. Insect Seminal Fluid Proteins: Identification and Function. *Annu Rev Entomol* **56**: 21–40.

Baer, B., Heazlewood, J.L., Taylor, N.L., Eubel, H. & Millar, A.H. 2009. The seminal fluid proteome of the honeybee Apis mellifera. *Proteomics* **9**: 2085–2097.

Bakst, M.R. 2011. PHYSIOLOGY AND ENDOCRINOLOGY SYMPOSIUM: Role of the oviduct in maintaining sustained fertility in hens. *Journal of Animal Science* **89**: 1323–1329.

Begun, D.J., Lindfors, H.A., Kern, A.D. & Jones, C.D. 2006a. Evidence for de novo evolution of testis-expressed genes in the *Drosophila yakuba*/*Drosophila erecta* clade. *Genetics* **176**: 1131–1137.

Begun, D.J., Lindfors, H.A., Thompson, M.E. & Holloway, A.K. 2006b. Recently evolved genes identified from *Drosophila yakuba* and *D. erecta* accessory gland expressed sequence tags. *Genetics* **172**: 1675–1681.

Benjamini, Y. & Hochberg, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 289–300. JSTOR.

Birkhead, T.R., Møller, A.P. & Sutherland, W.J. 1993. Why do females make it so difficult for males to fertilize their eggs? *J Theor Biol* **161**: 51–60. Elsevier.

Bourgeon, F. 2003. Involvement of Semenogelin-Derived Peptides in the Antibacterial

Activity of Human Seminal Plasma. *Biol Reprod* **70**: 768–774.

Bourin, M., Gautron, J., Berges, M., Attucci, S., Le Blay, G., Labas, V., *et al.* 2011. Antimicrobial Potential of Egg Yolk Ovoinhibitor, a Multidomain Kazal-like Inhibitor of Chicken Egg. *Journal of Agricultural and Food Chemistry* **59**: 12368–12374.

Buxton, J.R. & Orcutt, F.S. 1975. Enzymes and electrolytes in the semen of Japanese quail. *Poultry Sci* **54**: 1556–1566.

Callewaert, L. & Michiels, C.W. 2010. Lysozymes in the animal kingdom. *J Biosci* **35**: 127–160.

Carroll, S.B. 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**: 25–36.

Chapman, T., Liddle, L., Kalb, J., Wolfner, M.F. & Partridge, L. 1995. Cost of mating in *Drosophila melanogaster* females is mediated by male accessory-gland products. *Nature* **373**: 241–244.

Chapman, T., Neubaum, D.M., Wolfner, M.F. & Partridge, L. 2000. The role of male accessory gland protein Acp36DE in sperm competition in *Drosophila melanogaster*. *Proceedings Of The Royal Society B-Biological Sciences* **267**: 1097–1105.

Chen, F., Mackey, A.J., Stoeckert, C.J. & Roos, D.S. 2006. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Research* **34**: D363–8.

Cheng, K.M., Hickman, A.R. & Nichols, C.R. 1989. Role of the proctodeal gland foam of male Japanese quail in natural copulations. *Auk* **106**: 279–285.

Cheng, K.M., McIntyre, R.F. & Hickman, A.R. 1989. Proctodeal gland foam enhances competitive fertilization in domestic Japanese quail. *Auk* **106**: 286–291.

Civetta, A. & Singh, R.S. 1995. High divergence of reproductive tract proteins and their association with postzygotic reproductive isolation in Drosophila melanogaster and Drosophila virilis group species. *J Mol Evol* **41**: 1085–1095.

Clark, A.G., Aguade, M., Prout, T., Harshman, L.G. & Langley, C.H. 1995. Variation in sperm displacement and its association with accessory gland protein loci in *Drosophila melanogaster*. *Genetics* **139**: 189–201.

Clark, N.L., Aagaard, J.E. & Swanson, W.J. 2006. Evolution of reproductive proteins from animals and plants. *Journal of Reproduction and Fertility* **131**: 11–22.

Coil, W.H. & Wetherbee, D.K. 1959. Observations on the cloacal gland of the Eurasian quail *Coturnix coturnix*. *The Ohio Journal of Science* **59**: 268–270.

Coyne, J. & Orr, H. 2004. *Speciation*. Sinauer Associates, Inc. Sunderland, Massachusetts.

Das, S.C., Isobe, N. & Yoshimura, Y. 2008. REVIEW ARTICLE: Mechanism of Prolonged Sperm Storage and Sperm Survivability in Hen Oviduct: A Review. *American Journal of Reproductive Immunology* **60**: 477–481.

Dean, M.D. 2013. Genetic disruption of the copulatory plug in mice leads to severely reduced fertility. *Plos Genetics* **9**: e1003185.

Dean, M.D., Clark, N.L., Findlay, G.D., Karn, R.C., Yi, X., Swanson, W.J., *et al.* 2009. Proteomics and comparative genomic investigations reveal heterogeneity in evolutionary rate of male reproductive proteins in mice (*Mus domesticus*). *Molecular Biology and Evolution* **26**: 1733–1743.

Dean, M.D., Findlay, G.D., Hoopmann, M.R., Wu, C.C., MacCoss, M.J., Swanson, W.J., *et al.* 2011. Identification of ejaculated proteins in the house mouse (*Mus domesticus*) via isotopic labeling. *BMC Genomics* **12**: 306.

Dorus, S., Busby, S.A., Gerike, U., Shabanowitz, J., Hunt, D.F. & Karr, T.L. 2006. Genomic and functional evolution of the Drosophila melanogaster sperm proteome. *Nat Genet* **38**: 1440–1445.

Dorus, S., Wasbrough, E.R., Busby, J., Wilkin, E.C. & Karr, T.L. 2010. Sperm Proteomics Reveals Intensified Selection on Mouse Sperm Membrane and Acrosome Genes. *Molecular Biology and Evolution* **27**: 1235–1246.

Findlay, G.D. & Swanson, W.J. 2010. Proteomics enhances evolutionary and functional analysis of reproductive proteins. *BioEssays* **32**: 26–36.

Findlay, G.D., Maccoss, M.J. & Swanson, W.J. 2009. Proteomic discovery of previously unannotated, rapidly evolving seminal fluid genes in Drosophila. *Genome Research* **19**: 886–896.

Findlay, G.D., Yi, X., MacCoss, M.J. & Swanson, W.J. 2008. Proteomics reveals novel Drosophila seminal fluid proteins transferred at mating. *Plos Biol* **6**: e178. Public Library of Science.

Finseth, F.R., Bondra, E.R., & Harrison, R.G. 2013a. Phenotypic divergence arose without major protein-coding sequence divergence in a novel reproductive gland. Chapter 3, Ph.D. Thesis, Cornell University.

Finseth, F.R., Iacovelli, S.R., Harrison, R.G., & Adkins-Regan, E.K. 2013b. A non-semen copulatory fluid influences the outcome of sperm competition in Japanese quail. *Journal of Evolutionary Biology, in press.*

Ford, W.C.L. 2006. Glycolysis and sperm motility: does a spoonful of sugar help the flagellum go round? *Hum Reprod Update* **12**: 269–274.

Froman, D.P. & Feltmann, A.J. 1998. Sperm mobility: a quantitative trait of the domestic

fowl (*Gallus domesticus*). *Biol Reprod* **58**: 379–384.

Fujihara, N. 1992. Accessory reproductive fluids and organs in male domestic birds. *Worlds Poultry Science Journal* **48**: 39–56.

Higaki, K., Yoshimura, Y. & Tamura, T. 1995. Localization of Spermatozoa and Leukocytes in Vagina and Uterovaginal Junction after Copulation in Japanese Quail (Cotumix cotumix japónica).

Huang, D.W., Sherman, B.T. & Lempicki, R.A. 2009. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research* **37**: 1–13.

Huang, D.W., Sherman, B.T. & Lempicki, R.A. 2008. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**: 44–57.

Ibrahim, H.R., Sugimoto, Y. & Aoki, T. 2000. Ovotransferrin antimicrobial peptide (OTAP-92) kills bacteria through a membrane damage mechanism. *Biochim. Biophys. Acta* **1523**: 196–205.

Ikeda, K. & Taji, K. 1954. On the foamy ejaculate of Japanese quail, *Coturnix coturnix japonica*. *Scientific reports of the Matsuyama agricultural college* **13**: 1–4.

Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., *et al.* 2005. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol. Cell Proteomics* **4**: 1265–1272.

James, K. & Hargreave, T.B. 1984. Immunosuppression by seminal plasma and its possible clinical significance. *Immunol Today* **5**: 357–363. Elsevier.

Kaur, S., Prabha, V. & Shukla, G. 2010. Interference of Human Spermatozoal Motility by Live Staphylococcus aureus. *Am. J. Biomed. Sci.* 91–97.

Kelleher, E.S., Watts, T.D., Laflamme, B.A., Haynes, P.A. & Markow, T.A. 2009. Proteomic analysis of Drosophila mojavensis male accessory glands suggests novel classes of seminal fluid proteins. *Insect Biochemistry and Molecular Biology* **39**: 366–371. Elsevier Ltd.

King, A.S. 1981. Cloaca. In: *Form and function in birds* (A. S. King & J. McLelland, eds), pp. 63–105. Academic Press, New York.

Klemm, R., Knight, C.E. & Stein, S. 1973. Gross and microscopic morphology of glandula-proctodealis (foam gland) of *Coturnix c. japonica* (aves). *J Morphol* **141**: 171–184.

Knox, K. & Baker, J.C. 2008. Genomic evolution of the placenta using co-option and duplication and divergence. *Genome Research* **18**: 695–705.

Korn, N., Thurston, R., Pooser, B. & Scott, T. 2000. Ultrastructure of spermatozoa from Japanese quail. *Poultry Sci* **79**: 407–414.

Lake, P. 1981. Male genital organs. In: *Form and function in birds* (A. S. King & J. McLelland, eds), pp. 1–61. Academic Press, London.

Li, H. & Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., *et al.* 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Long, M., Betrán, E., Thornton, K. & Wang, W. 2003. The origin of new genes: glimpses from the young and old. *Nature Reviews Genetics* **4**: 865–875.

Mann, K. 2007. The chicken egg white proteome. *Proteomics* **7**: 3558–3568.

McFarland, L. Z., Warner, R.L., Wilson, W.O. & Mather, F.B. 1968. Cloacal gland complex of japanese quail. *Experientia* **24**: 941–&.

Min, X.J., Butler, G., Storms, R. & Tsang, A. 2005. OrfPredictor: predicting protein-coding regions in EST-derived sequences. *Nucleic Acids Research* **33**: W677–W680.

Moretti, E., Capitani, S., Figura, N., Pammolli, A., Federico, M.G., Giannerini, V., *et al.* 2008. The presence of bacteria species in semen and sperm quality. *J Assist Reprod Genet* **26**: 47–56.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. & Wold, B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Meth* **5**: 621–628.

Mueller, J., Ram, K.R., McGraw, L., Qazi, M., Siggia, E., Clark, A., *et al.* 2005. Cross-species comparison of Drosophila male accessory gland protein genes. *Genetics* **171**: 131–143.

Nakano, T. & Graf, T. 1991. Goose-type lysozyme gene of the chicken: sequence, genomic organization and expression reveals major differences to chicken-type lysozyme gene. *Biochim. Biophys. Acta* **1090**: 273–276.

O'Neil, S.T., Dzurisin, J.D., Carmichael, R.D., Lobo, N.F., Emrich, S.J. & Hellmann, J.J. 2010. Population-level transcriptome sequencing of nonmodel organisms Erynnis propertius and Papilio zelicaon. *BMC Genomics* **11**: 310.

Otti, O., McTighe, A.P. & Reinhardt, K. 2012. In vitroantimicrobial sperm protection by an ejaculate-like substance. *Functional Ecology* **27**: 219–226.

Pellegrini, A., Hülsmeier, A.J., Hunziker, P. & Thomas, U. 2004. Proteolytic fragments of ovalbumin display antimicrobial activity. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1672**: 76–85.

Petersen, T.N., Brunak, S., Heijne, von, G. & Nielsen, H. 2011. nmeth.1701. *Nat Meth* **8**: 785–786. Nature Publishing Group.

Pitnick, S., Wolfner, M.F. & Suarez, S.S. 2008. *Ejaculate–female and sperm–female interactions* (T. R. Birkhead, D. J. Hosken, & S. Pitnick, eds). Academic Press, London.

Poiani, A. 2006. Complexity of seminal fluid: a review. *Behav Ecol Sociobiol* **60**: 289–310.

Poiani, A. 2002. Sperm competition promoted by sexually transmitted pathogens and female immune defences. *Ethol Ecol Evol* **14**: 327–340.

Prager, E.M., Wilson, A.C. & Arnheim, N. 1974. Widespread distribution of lysozyme g in egg white of birds. *Journal of Biological Chemistry* **249**: 7295–7297. ASBMB.

Ram, K.R. & Wolfner, M.F. 2007. Seminal influences: *Drosophila* Acps and the molecular interplay between males and females during reproduction. *Integr Comp Biol* **47**: 427–445.

Ramm, S.A., McDonald, L., Hurst, J.L., Beynon, R.J. & Stockley, P. 2009. Comparative proteomics reveals evidence for evolutionary diversification of rodent seminal fluid and its functional significance in sperm competition. *Molecular Biology and Evolution* **26**: 189–198.

Rice, W. 1998. Intergenomic conflict, interlocus antagonistic coevolution, and the evolution of reproductive isolation. In: *Endless forms: species and speciation* (D. J. Howard & S. H. Berlocher, eds), pp. 261–270. Oxford Univ Press, New York.

Robertson, S.A. 2007. Seminal fluid signaling in the female reproductive tract: Lessons from rodents and pigs. *Journal of Animal Science* **85**: E36–E44.

Robertson, S.A. 2005. Seminal plasma and male factor signalling in the female reproductive tract. *Cell Tissue Res* **322**: 43–52.

Robertson, S.A., Guerin, L.R., Moldenhauer, L.M. & Hayball, J.D. 2009. Activating T regulatory cells for tolerance in early pregnancy — the contribution of seminal fluid. *J. Reprod. Immunol.* **83**: 109–116.

Robertson, S.A., Ingman, W.V., O'Leary, S., Sharkey, D.J. & Tremellen, K.P. 2002. Transforming growth factor beta--a mediator of immune deviation in seminal plasma. *J. Reprod. Immunol.* **57**: 109–128.

Robinson, M.D. & Oshlack, A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology* **11**: R25.

Robinson, M.D., McCarthy, D.J. & Smyth, G.K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.

Rowe, M., Czirjak, G.A. & Lifjeld, J.T. 2013. Lysozyme-associated bactericidal activity in the ejaculate of a wild passerine. *Biological Journal of …*.

Rowe, M., Czirjak, G.A., McGraw, K.J. & Giraudeau, M. 2011. Sexual ornamentation reflects antibacterial activity of ejaculates in mallards. *Biology Letters* **7**: 740–742.

Sachs, B.D. 1967. Photoperiodic control of cloacal gland of japanese quail. *Science* **157**: 201–&.

Schumacher, M. & Balthazart, J. 1983. The effects of testosterone and its metabolites on sexual-behavior and morphology in male and female Japanese quail. *Physiol Behav* **30**: 335–339.

Seiwert, C. & Adkins-Regan, E. 1998. The foam production system of the male Japanese quail: Characterization of structure and function. *Brain Behav Evol* **52**: 61–80.

Sheldon, B.C. 1993. Sexually transmitted disease in birds: occurrence and evolutionary significance. *Philos T Roy Soc B* **339**: 491–497.

Singh, R., Sastry, K., Pandey, N. & Shit, N. 2011. Molecular Characterization and Expression of LDHA and LDHB mRNA in Testes of Japanese Quail (Coturnix japonica).

Singh, R.P., H Sastry, von, K., Shit, N., Pandey, N.K., Singh, K.B., Mohan, J., *et al.* 2011a. Cloacal gland foam enhances motility and disaggregation of spermatozoa in Japanese quail (*Coturnix japonica*). *Theriogenology* **75**: 563–569. Elsevier Inc.

Singh, R.P., Sastry, K.V.H., Pandey, N.K., Shit, N., Agrawal, R., Singh, K.B., *et al.* 2011b. Characterization of lactate dehydrogenase enzyme in seminal plasma of Japanese quail (Coturnix coturnix japonica). *THE* **75**: 555–562. Elsevier Inc.

Singh, R.P., Sastry, K.V.H., Pandey, N.K., Singh, K.B., Malecki, I.A., Farooq, U., *et al.* 2012. The role of the male cloacal gland in reproductive success in Japanese quail (*Coturnix japonica*). *Reprod. Fertil. Dev.* **24**: 405.

Stockley, P. 1997. Sexual conflict resulting from adaptations to sperm competition. *Trends Ecol Evol*.

Swanson, W. & Vacquier, V. 2002. Reproductive protein evolution. *Annual Review of Ecology and Systematics* **33**: 161–179.

Swanson, W., Clark, A., Waldrip-Dail, H., Wolfner, M.F. & Aquadro, C. 2001a. Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in Drosophila. *P Natl Acad Sci Usa* **98**: 7375–7379.

Swanson, W., Zhang, Z., Wolfner, M.F. & Aquadro, C. 2001b. Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. *P Natl Acad Sci Usa* **98**: 2509–2514.

True, J. & Carroll, S. 2002. Gene co-option in physiological and morphological evolution. *Annual Review of Cell and Developmental Biology* **18**: 53–80.

Turner, L.M. & Hoekstra, H.E. 2008. Causes and consequences of the evolution of reproductive proteins. *Int. J. Dev. Biol.* **52**: 769–780.

Turner, R.M. 2003. Tales from the tail: what do we really know about sperm motility? *Journal Of Andrology* **24**: 790–803.

Van Belleghem, S.M., Roelofs, D., Van Houdt, J. & Hendrickx, F. 2012. De novo transcriptome assembly and SNP discovery in the wing polymorphic salt marsh beetle Pogonus chalceus (Coleoptera, Carabidae). *PLoS ONE* **7**: e42605.

Visconti, P.E. 2012. Sperm Bioenergetics in a Nutshell. *Biol Reprod* **87**: 72–72.

Wagstaff, B.J. 2005. Molecular population genetics of accessory gland protein genes and testis-expressed genes in *Drosophila mojavensis* and *D. arizonae*. *Genetics* **171**: 1083–1101.

Wagstaff, B.J. & Begun, D.J. 2007. Adaptive evolution of recently duplicated accessory gland protein genes in desert *Drosophila*. *Genetics* **177**: 1023–1030.

Wahl, S.M., Hunt, D.A., Wong, H.L., Dougherty, S., McCartney-Francis, N., Wahl, L.M., *et al.* 1988. Transforming growth factor-beta is a potent immunosuppressive agent that inhibits IL-1-dependent lymphocyte proliferation. *The Journal of Immunology* **140**: 3026–3032. Am Assoc Immnol.

Walters, J.R. & Harrison, R.G. 2010. Combined EST and Proteomic Analysis Identifies Rapidly Evolving Seminal Fluid Proteins in Heliconius Butterflies. *Molecular Biology and Evolution* **27**: 2000–2013.

Weinberg, E.D. 2001. Human lactoferrin: a novel therapeutic with broad spectrum potential. *J. Pharm. Pharmacol.* **53**: 1303–1310.

Wesierska, E., Saleh, Y., Trziszka, T., Kopec, W., Siewinski, M. & Korzekwa, K. 2005. Antimicrobial activity of chicken egg white cystatin. *World J Microbiol Biotechnol* **21**: 59–64.

Wishart, G.J. 1982. Maintenance of ATP concentrations in and of fertilizing ability of fowl and turkey spermatozoa in vitro. *Journal of Reproduction and Fertility* **66**: 457–462. Soc Reprod Fertility.

Wishart, G.J. 1981. The effect of continuous aeration on the fertility of fowl and Turkey semen stored above 0 °C. *British Poultry Science* **22**: 445–450.

Wolfner, M.F. & Wolfner, M.F. 2009. Battle and Ballet: Molecular Interactions between the Sexes in Drosophila. *Journal Of Heredity* **100**: 399–410.

Woolley, D.M. 1995. The structure of the spermatozoon of the Japanese quail, Coturnix coturnix L., var. japonica. *Acta Zoologica* **76**: 45–50. Wiley Online Library.

Wray, G.A. 2007. The evolutionary significance of cis-regulatory mutations. *Nature Reviews Genetics* **8**: 206–216.

Zhao, S. & Fernald, R.D. 2005. Comprehensive algorithm for quantitative real-time polymerase chain reaction. *Journal of Computational Biology* **12**: 1047–1064.

APPENDIX

SUPPLEMENTARY INFORMATION

Chapter 1

The following are supplementary tables for Chapter 1.

Table S1.1 Primers, size range, number of alleles, fluorescent tags, primer mixes, number of adults genotyped, observed and expected heterozygosities, polymorphic information content (PIC), P values and frequency of null alleles based on 88 adult quail. GUJ0041 was not used for paternity analyses due to high incidence of null alleles. All primers were originally described by Kayang et al., (2000, 2002). Details including primer sequence, repeat array, and accession numbers can be found in those references.

Primer	Size Range	# of alleles	Tag	Mix	*N*	H_o	H_e	PIC	*P* value	Null frequency
GUJ0029	167-183	4	FAM	1	88	0.716	0.736	0.683	0.376	0.014
GUJ0041	139-153	4	NED	2	88	0.303	0.683	0.622	0.000	0.382
GUJ0044	207-245	5	NED	2	88	0.625	0.664	0.601	0.095	0.029
GUJ0049	256-265	5	NED	2	88	0.557	0.634	0.563	0.111	0.071
GUJ0052	121-129	4	FAM	1	88	0.659	0.674	0.623	0.906	0.009
GUJ0068	230-264	6	FAM	1	88	0.568	0.558	0.522	0.942	-0.004

Table S1.2 Parameter estimates and standard errors (SE) of best models for (a) proportion of a clutch fertilized by a non-manipulated male, probability of a (b) non-manipulated and (c) manipulated male fertilizing any eggs, and (d) duration of fertility. Estimates were generated by maximum likelihood for (a)-(c) and restricted maximum likelihood for (d). Estimates from significant effects are in bold.

Best model term	k[§]	Estimate[¶]	SE[¶]
a) Proportion of clutch fertilized by a non-manipulated male			
	5		
Intercept		0.581	0.60
Foam = Present		**0.243**[‡]	**0.57**
Mating order = Non-manipulated first		**0.827**[‡]	**0.57**
b) Probability of a non-manipulated male fertilizing any eggs			
	4		
Intercept		0.641	0.62
Foam = Present		0.299	0.63
Mating order = Non-manipulated first		**0.763**[*]	**0.63**
c) Probability of a manipulated male fertilizing any eggs			
	4		
Intercept		0.281	0.59
Foam = Present		**0.803**[†]	**0.62**
d) Duration of fertility			
	6		
Intercept		0.997	0.61

[*] $P < 0.05$
[†] $P < 0.005$
[‡] $P < 0.001$
[§] k represents number of parameters.
[¶] Parameter estimates and standard errors are on raw scale.

Table S1.3: Means and/or standard errors (SE) of the (a) proportion of clutches fertilized by a non-manipulated male and (d) last days a manipulated male fertilized an egg. Proportions of trials where a (b) non-manipulated and (c) manipulated male fertilized any eggs. Overall sample sizes given in (N).

Manipulated male position	Foam	Overall	Group 1	Group 2	N
a) Mean proportions of clutches fertilized by non manipulated males		Mean (SE)	Mean (SE)	Mean (SE)	
1	Absent	0.65 (0.12)	0.57 (0.51)	0.75 (0.43)	21
1	Present	0.33 (0.95)	0.31 (0.41)	0.40 (0.46)	22
2	Absent	0.78 (0.08)	0.83 (0.33)	0.72 (0.34)	20
2	Present	0.56 (0.10)	0.53 (0.45)	0.60 (0.39)	20
b) Proportion of trials a non-manipulated male fertilized any eggs		Proportion	Proportion	Proportion	
1	Absent	0.57	0.55	0.60	21
1	Present	0.50	0.62	0.33	22
2	Absent	0.85	0.82	0.89	20
2	Present	0.60	0.46	0.78	20
d) Proportion of trials a manipulated male fertilized any eggs		Proportion	Proportion	Proportion	
1	Absent	0.29	0.36	0.20	21
1	Present	0.68	0.77	0.56	22
2	Absent	0.35	0.27	0.44	20
2	Present	0.55	0.55	0.56	20
d) Means of the last day a manipulated male fertilized an egg		Mean (SE)	Mean (SE)	Mean (SE)	
1	Absent	6.33 (0.62)	6.25 (0.95)	6.50 (0.50)	6
1	Present	5.60 (0.57)	5.40 (0.82)	6.00 (0.55)	15
2	Absent	5.57 (1.00)	5.33 (2.19)	5.75 (1.03)	7
2	Present	5.63 (0.47)	5.83 (0.40)	5.40 (0.98)	11

Chapter 2

The following is a supplementary table for Chapter 2.

Table 2.1: Summary statistics of raw data

Summary Statistic	454	Illumina
Total number of sequences	383,803	220,127,867
Average length	222 (40-696)	100
Median length	217	100
Sequences suitable for assembly	266,097	135,907,791
Approximate cost of sequencing	$6,000	$4,500

Chapter 3

The following are supplementary methods, tables, and figures for Chapter 3.

Methods S3.1

Birds and sequencing

Japanese quail were lab-reared and housed on a 16:8 light:dark cycle to simulate breeding conditions. Three tissue samples (foam gland, testis, and liver) were collected from six sexually mature Japanese quail males with phenotypically normal foam glands. Two males were approximately one-year old and euthanized in February 2011. The remaining four males were approximately five months old and euthanized in November 2011. After euthanizing with CO_2, we immediately dissected out foam gland, testis, and liver tissues from each male and froze samples on liquid nitrogen. Tissues were later frozen at -80 °C until RNA extraction.

Library preparation and sequencing

We extracted RNA from the 18 samples (3 tissues x 6 males) using the Agencourt® RNAdvance™ Tissue Kit (Beckman Coulter) and following the manufacturer's instructions except that we performed half-reactions. RNA quality and concentration was assessed by agarose gel electrophoresis and NanoDrop™ spectrophotometry. We confirmed RNA purity and integrity using an Agilent 2100 BioAnalyzer. In January 2012, eighteen Illumina libraries were prepared from 1.2 μg total RNA using the TruSeq™ RNA Sample Preparation Kit (Illumina®) following manufacturer's instructions. All 18 samples were tagged with a unique adapter index, pooled, and single-end sequenced on the equivalent of two lanes of

an Illumina HiSeq 2000, with a target read length of 100 bp. Sequencing was performed by the Cornell University Institute of Biotechnology's Genomics Facility.

Transcriptome assembly and characterization

Initial quality filtering and barcode removal were conducted by the Cornell University Institute of Biotechnology's Genomics Facility. Unless otherwise stated, bioinformatics analyses were conducted on a Linux, Dell Precision T3500n with 4 cores, 24 GB RAM, and 4 TB HDD housed at the Cornell University's Institute of Biotechnology's Bioinformatics Facility. We used fastq-mcf (http://code.google.com/p/ea-utils/wiki/FastqMcf) to remove Illumina adaptors, trim low-quality terminal ends, discard short sequences, and filter reads with phred scores < 20. We merged the 18 libraries into a single file and assembled a transcriptome using the Trinity pipeline with default parameters (Grabherr *et al.*, 2011). Trinity previously produced high-quality transcriptomes via similar protocols for this non-model species (Finseth & Harrison 2013). The Trinity assembly was executed on a Linux, Dell PowerEdge R710 with 16 cores, 64 GB RAM, and 1 TB HDD also housed at Cornell's Bioinformatics Facility. The resulting library had 85,900 contigs. We reduced all contigs with putative isoforms identified by Trinity (*i.e.*, multiple contigs with same component number) to the single longest isoform (removed 2,677 contigs). In 27 instances, there was not a single longest contig. In these cases, we retained the isoform with the best BLASTp to the chicken, or, if that did not break the tie, we kept the first sequence generated by Trinity. We identified open reading frames with OrfPredictor (Min *et al.*, 2005) and kept only those sequences with at least one open

reading frame. The transcriptome had 81,868 contigs and is referred to as the 'exhaustive' transcriptome.

Filtered reads from each sample were aligned to the transcriptome using the Burrow-Wheeler transform in BWA (H. Li & Durbin, 2009). The number of reads per sample uniquely mapped to each contig was tabulated with samtools as implemented in custom python scripts (H. Li *et al.*, 2009). To generate a set of high-quality contigs that represent real transcripts, we filtered out very lowly expressed contigs by retaining only those contigs with at least one read aligned per every million reads for at least six samples (i.e., the number of biological replicates per tissue). We refer to the remaining 24,035 transcripts as the 'filtered' transcriptome throughout. We calculated standard metrics of transcriptome quality for the filtered transcriptome including: average transcript length, median transcript length, N50 (median transcript size, weighted by length), and summed transcript lengths (Kumar & Blaxter, 2010; Hornett & Wheat, 2012).

To evaluate whether our assembly returned full-length transcripts, we computed the "ortholog hit ratio" for each transcript from the filtered transcriptome, as described by O'Neil *et al*., (2010). Ortholog assignment is described in Section 2.4. This ratio represents the length of a putative coding region of a quail transcript divided by the length of the coding region of the orthologous chicken transcript. For calculation of ortholog hit ratios, the putative coding region was estimated from the alignment length of the best BLASTp result between a transcript and its chicken ortholog. An ortholog completely represented by a transcript would have a ratio of "1". Ratios less than 1 indicate instances where transcripts only partially covered orthologs, while ratios greater than 1 indicate insertions in transcripts.

Real-time quantitative PCR validation of RNA-Seq

To validate RNA-Seq, we performed RT-qPCR on the 18 RNA samples used for RNA-Seq (6 replicates x 3 tissues). RNA was treated with Turbo™ DNase (Ambion®) and confirmed to be free of genomic DNA by attempting to PCR amplify a panel of three housekeeping genes (see Table S#) directly from RNA. Two hundred ng of RNA was reverse transcribed into cDNA with SuperScript® III First Strand cDNA Synthesis Kit (Invitrogen™) following the manufacturer's instructions. Primers were designed from nine genes significantly 'enriched' in one of the three tissues (three from each tissue). We verified that primers amplified the intended target by Sanger sequencing. β-actin served as an internal control and was confirmed to be stable across treatments using Normfinder (Andersen *et al.*, 2004). RT-qPCR reactions (25 μL) were performed in duplicate with 33 ng of cDNA template and 200 nM of each primer using the Power SYBR® Green Master Mix (Applied Biosystems). Samples were run on a ViiA™ 7 (Applied Biosystems) thermocycler with the following parameters: 95° C for 10 min, 40 cycles of 95° C for 15s and 60° C for 60s. Primer efficiencies were calculated with Real-time PCRMiner and ranged from 95 - 100% (Zhao & Fernald, 2005). We followed the $2^{-\Delta CT}$ method as described by Schmittgen *et al*., (2008). We generated 95% confidence intervals around these values for each tissue to ascertain whether genes were significantly enriched in the expected tissues. Primer sequences, target sizes, efficiencies, and log fold changes are given in Table S3.3.

Table S3.1 Average and standard deviations for sequencing reads.

Type	Average	Standard Deviation
Raw	22072077	8810376
Filtered	11909891	6010577
Mapped	5365981	3067257
Uniquely mapped	4922004	2806453

Table S3.2 Standard metrics of transcriptome quality

Metric	Filtered Transcriptome
Number of transcripts	24,035
Average transcript length	1687.544
Median transcript length	1183
N50	552
Sum of transcript lengths	40,560,131

Table S3.3 RT-qPCR validation of RNA-Seq data

Gene	Treatment[a]	Forward primer (5' - 3')	Reverse primer (5' - 3')	Amplicon size (bp)	Avg efficiency[b]	SD efficiency[b]	FG (95% CI)[c]	Testis (95% CI)[c]	Liver (95%CI)[c]
Coja4716	Foam gland	TTTAAGGCAACTGACCATCC	CTCTCACTGTTTCTTCCATCC	167	0.9692	0.0083	0.650 (0.403 - 0.865)	0.001 (0.001 - 0.002)	0.001 (0.000 - 0.002)
Coja12766	Liver	GATGTGAAGGATGAAGTCAGG	AGCCATAGAGACTTGTAGGG	172	0.9839	0.0063	0.000 (0.000 - 0.000)	0.001 (0.001 - 0.002)	0.277 (0.127 - 0.446)
Coja4787	Liver	TTTAAGGCAACTGACCATCC	CTCTCACTGTTTCTTCCATCC	167	0.9682	0.0132	0.000 (0.000 - 0.001)	0.000 (0.000 - 0.000)	0.420 (0.238 - 0.613)
Coja13961	Liver	GAGCTGTGAAATAGCTGTGG	CACACCATCACTTGCTAAGG	139	1.0011	0.0046	0.000 (0.000 - 0.000)	0.000 (0.000 - 0.000)	0.060 (0.023 - 0.096)
Coja12001	Foam gland	ACACCGATTTCCCATTTCC	TGCTTATAACGCCGATGC	141	0.9743	0.0027	0.236 (0.102 - 0.385)	0.000 (0.000 - 0.001)	0.001 (0.000 - 0.001)
Coja12871	Foam gland	TCATCTCCATCCCTGTATCC	GCTGCTGAGAAACAGACC	158	0.9512	0.0077	1.102 (0.731 - 1.491)	0.003 (0.002 - 0.004)	0.001 (0.000 - 0.013)
Coja12292	Testis	GAGCTGGATGAGATGATTGG	CCTTGAAGAGCTGTTTCTCC	178	0.9770	0.0079	0.000 (0.000 - 0.000)	1.346 (0.659 - 1.989)	0.000 (0.000 - 0.001)
Coja15009	Testis	GAGGCTACGGCTTATTTGG	CTGGCTGTTTCTTTCTTTGC	166	0.9976	0.0067	No amplification	0.161 (0.089 - 0.232)	0.000 (0.000 - 0.000)
Coja20376	Testis	CCACTTCACACCAACTTCC	AAACCTTGCGTTCTTGGG	149	0.9759	0.0045	0.000 (0.000 - 0.000)	0.829 (0.736 - 0.961)	0.000 (0.000 - 0.000)
β-actin	Internal control/ Housekeeping	ATGCAGAAGGAGATCACAGC	TGTTGGTAACAGTCCGGTTT	200	0.9579	0.0041	-	-	-
Cyclophilin[d]	Housekeeping	CCCTGACGAGAACTTCAAGC	TTCTCCACCTTCCTCACCAC	200	0.9318	0.0052	-	-	-
Glyceraldehyde-3 phosphate dehydrogenase[e]	Housekeeping	GAAGCTTACTGGAATGGCTT	CAGACGGCAGGTCAGGTCA	~100	0.9633	0.0046	-	-	-

[a] Tissue-based treatments mean the gene was significantly enrched in the given tissue versus the average of the other two

[b] Average (Avg) and standard deviations (SD) of efficiencies calculated in Real-Time PCRMiner (Zhao & Fernald 2005)

[c] Mean $2^{-\Delta CT}$ values calculated according to Schmittgen *et al.,* 2008; 95% CIs generated from 10,000 bootstrapped resamplings of the means

[d] Designed by Foye-Jackson *et al.*, (2011)

[e] Designed by Charlier *et al.*, (2006)

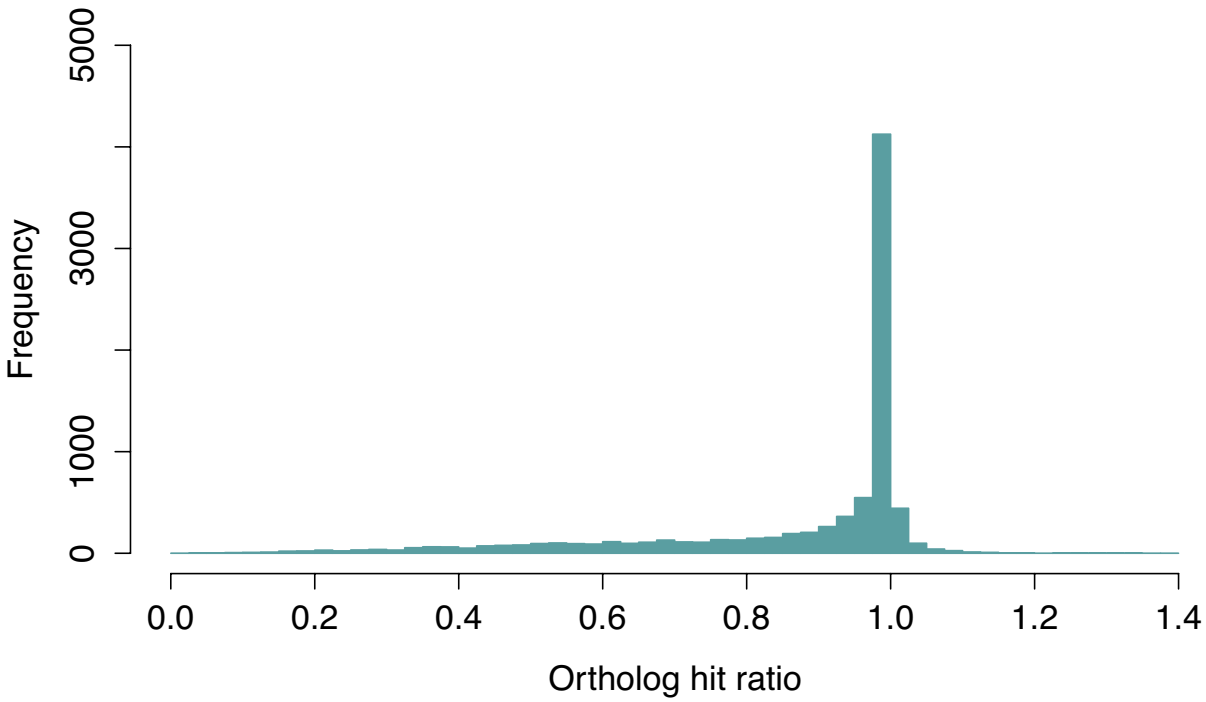

Figure S3.1. Histogram of ortholog hit ratios (quail transcript lengths relative to chicken ortholog length) for orthologs from the filtered transcriptome (N = 8,668). Ratios equal to 1 indicate fully assembled transcripts. Values < 1 signify partial transcripts, while values > 1 than represent contigs with insertions relative to orthologs.

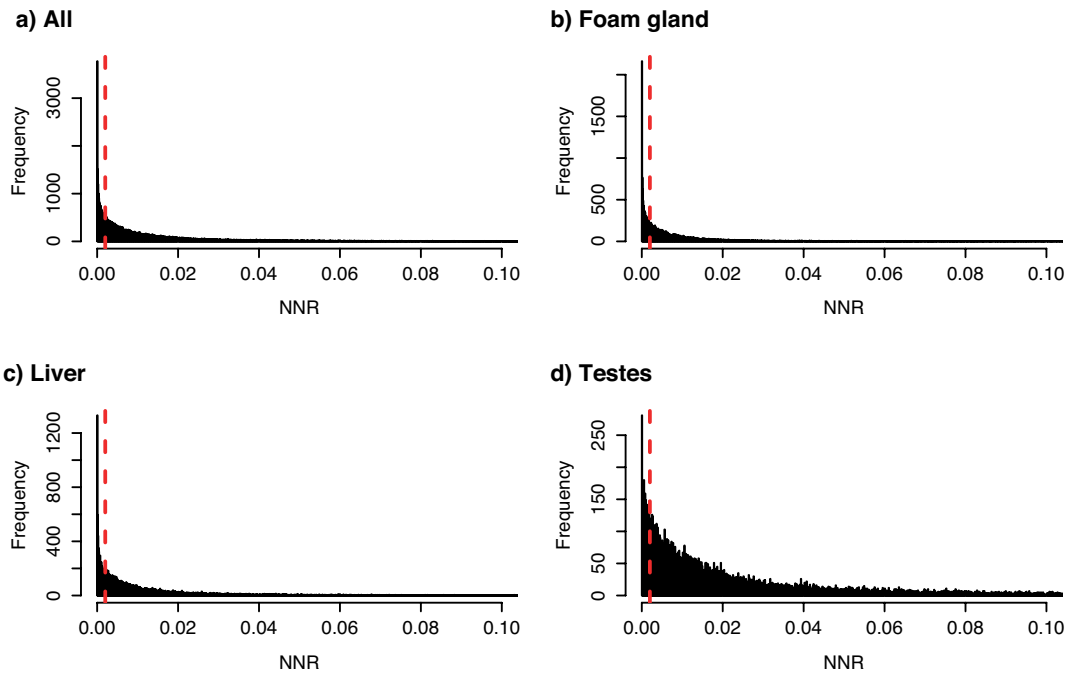

Figure S3.2 Normalized numbers of reads (NNR) for a) all transcripts from the exhaustive transcriptome and just those mapped reads from the b) foam gland, c) liver, and d) testis samples. Transcripts with NNR values to the right of the red dotted line in b-d were considered 'expressed' in the various tissues (*i.e.*, above noise). See Methods: '*Tissue-restricted'* for more details.

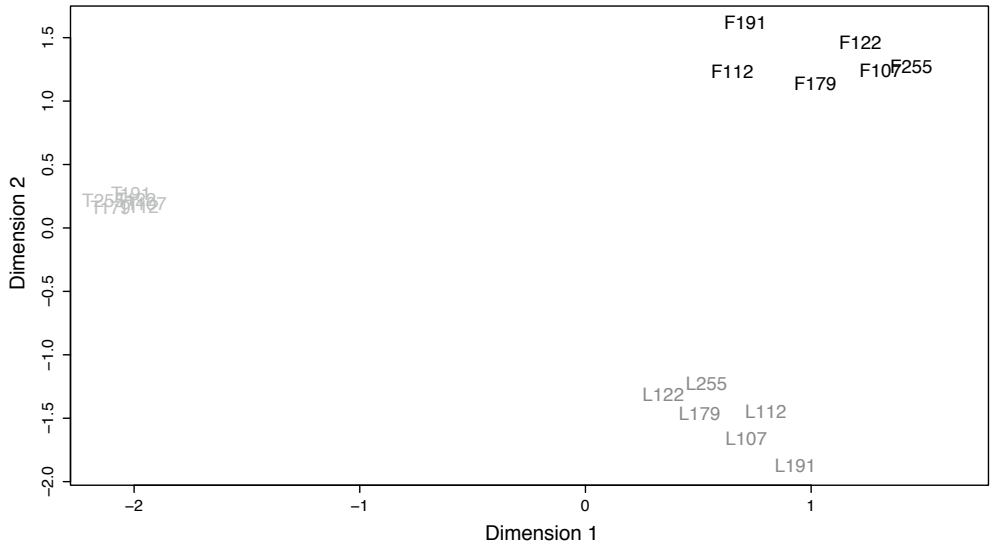

Figure S3.3. Multidimensional scaling plot of foam gland ("F", black), liver ("L", dark grey), and testis ("T", light grey) profiles from six male Japanese quail. Numbers after letters refer to male IDs. Pairwise distances correspond to the biological coefficient of variation or the coefficient of variation with which the (unknown) true abundance of the gene varies between RNA samples (McCarthy *et al.*, 2012). Thus, it represents the coefficient of variation that would remain between biological replicates if sequencing depth could be increased indefinitely.

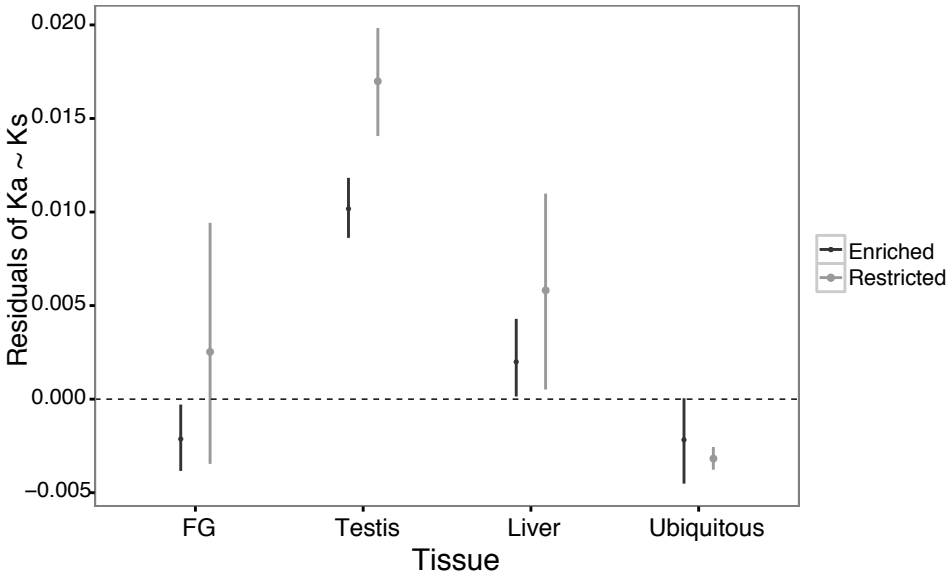

Figure S3.4. Average residuals (point) and 95% CIs (lines) from models regressing d_N on d_S across tissue classes for enriched (black) and restricted (grey) genes. Estimates are based on 1:1 orthologs between quail and chicken. Tissue designations are given in Figure 3.1. CIs derived from 10,000 bootstrap resamplings of the mean without assuming normality. Samples sizes as in Table 3.1. FG = Foam gland.

Chapter 4

The following are supplementary methods and tables for Chapter 4.

Methods S4.1. Protein preparation and mass spectrometry

Approximately equal volumes of foam were collected at 8 am from six sexually mature, 12-month old Japanese quail males and immediately prepared for SDS-PAGE protein separation. Each sample was suspended in 200 µL 1X phosphate buffered solution and centrifuged. The soluble fraction was transferred to a clean tube and left at room temperature for 4 hours to completely collapse the foam (air introduced during pipetting causes foam to reform). Equal parts of this solution from each male were pooled. We prepared, reduced, and denatured the pooled sample according to the NuPAGE® (Invitrogen™) instructions for denaturing gel electrophoresis under reducing conditions. Twenty µL of prepared sample were loaded onto a 4-12% NuPAGE® Novex Bis-Tris gel (Invitrogen™), separated with one-dimensional gel electrophoresis, and fixed according to the manufacturer's instructions. The resulting 1D SDS PAGE gel was stained with Coomassie Blue solution (0.025% Coomassie, 10% acetic acid), de-stained overnight, and kept at 4 °C in deionized water until preparation for MS/MS two days later. Care was taken to avoid contamination of the gel with human proteins.

In-gel digestion, tryptic peptide extractions, and nanoLC-MS/MS were conducted by the Cornell University Institute of Biolotechnology Proteomics and Mass Spectrometry Facility. One lane from the 1D SDS PAGE prepared above was cut into 8 bands. Each band was reduced with 10mM Dithiothreitol (Roche), alkylated with 55mM Iodoacetamide (Acros Organics), digested with trypsin (Promega™ sequencing grade), then extracted

twice with 50% acetonitrile (ACN), 5% Formic Acid (FA), and one final time with 90% ACN, 5% FA. The extracted peptides were lyophilized and reconstituted in 2% ACN, 0.5% FA for nanoLC-MS/MS analysis on an LTQ-Orbitrap Velos Mass Spectrometer (Thermo-Fisher Scientific, San Jose, CA) equipped with a CorConneX nano ion source (CorSolutions LLC, Ithaca, NY) and coupled to an UltiMate3000 nano Liquid Chromatograph (Dionex). Reconstituted peptides were injected onto a PepMap100 C18 trapping column (5µm particle size, 100Å pore size, 300µm inner diameter × 2cm length; Dionex) at 20µL/min for on-line desalting and separated on a PepMap100 C18 RP nano analytical column (3µm, 100Å, 75µm x 15cm; Dionex) installed on the CorConneX ion source with a 10µm emitter tip (NewObjective, Woburn, MA) and mounted in front of the instrument orifice. Peptides were eluted from the analytical column with a 60min gradient of 5 to 40% ACN in 0.1% formic acid at 300nL/min., followed by a 3min ramp to 95% ACN, 0.1%FA and a 5-min hold at 95% ACN, 0.1%FA. The column was then re-equilibrated with 2% ACN, 0.1%FA for 20mins prior to the next injection. The instrument was operated in positive polarity, with a nano spray voltage of 1.6 kV, an ion source capillary temperature of 275 °C, in data-dependent acquisition (DDA) mode using the FT mass analyzer for one survey (MS) scan of precursor ions followed by 10 MS/MS scans of the most intense ions with charge states +2 through +4 above a threshold ion count of 5,000 in the LTQ mass analyzer. MS survey scans were acquired at a resolution of 60,000 (fwhm at *m/z* 400) for the mass range of m/z 300-1800 and dynamic exclusion parameters were set at repeat count 1 with a 30s repeat duration, exclusion list size of 500, 15s exclusion duration, and ±10 *ppm* exclusion mass width. Collision induced dissociation (CID) parameters were set at the following values: isolation width 2.0 m/z, normalized collision energy 35 %, activation Q at 0.25, and an

activation time of 10 ms. Internal calibration using the background ion signal for polysiloxane at m/z 445.120025 as a lock mass in addition to external calibration of the FT mass analyzer was performed. All data were acquired using Xcalibur 2.1 software (Thermo-Fisher Scientific, San Jose, CA).

All MS and MS/MS spectra were processed using Proteome Discoverer 1.3 (Thermo-Fisher Scientific, San Jose, CA), and the raw data were exported as MGF files for subsequent database search using Mascot Daemon (version 2.3.02, Matrix Science, London, UK). The open reading frames were predicted for all transcripts from the exhaustive transcriptome (described in Finseth *et al.* 2013a; N = 81,868) with OrfPredictor (Min *et al.*, 2005). The acquired spectra were searched against these protein sequences with one missed cleavage by trypsin allowed. Peptide mass tolerance was set to 20 ppm and MS/MS mass tolerance was set to 0.8 Da. Carbamidomethylation of cysteine was set as a fixed modification, oxidation of methionine as well as deamidation of asparagine and glutamine were set as variable modifications. All matches at or above the 99% confidence threshold were considered confidently matched peptides. Proteins with at least 2 unique peptide matches, at least one of which matched only a single region in the transcriptome, constituted a preliminary list of 1006 potential genes encoding foam proteins (‘*P*’).

Table S4.1. RT-qPCR primers and validation of RNA-Seq data.

Gene	Treatment[a]	Forward primer (5' - 3')	Reverse primer (5' - 3')	Amplicon size (bp)	Avg efficiency[b]	SD efficiency[b]	logFC RNA-Seq (LD)	logFC RT-qPCR (LD)	logFC RNA-Seq (SD +T)	logFC RT-qPCR (SD + T)
Coja12001	Up-regulated	ACACCGATTTCCCATTTCC	TGCTTATAACGCCGATGC	141	0.993	0.004	3.594	2.790	3.508	1.985
Coja15808	Up-regulated	CTGGGCTGTTTGTTGTTATG	GGCCTTCTAGGCTGTATTG	149	0.957	0.006	3.620	3.175	3.770	2.782
Coja16773	Up-regulated	GATCCTGGAGACCCTTATCA	CATCCAGAGCCTTCACTTTC	158	0.968	0.007	2.015	0.710	1.914	1.184
Coja18519	Up-regulated	CCTCAAGCCAGCTCTAGTA	CTTGTCAGCCAGCATCTATC	161	1.009	0.004	2.328	2.307	2.308	2.009
Coja24228	Up-regulated	TCACCACAGGCACTATCA	GGCCTGTAGAAGGGTATCT	150	0.972	0.006	2.016	1.177	2.018	1.482
Coja26527	Up-regulated	CCGCATCAGTAAGGAGAAAG	TGAATGGCTCACACAAGAG	160	0.956	0.006	2.798	2.123	2.723	2.058
Coja43693	Up-regulated	GACGGGAAGAGGCAAATAC	AAGGCTCACGAACAAAGAG	152	0.986	0.004	6.084	4.650	6.336	4.275
Coja4537	Up-regulated	CCTACTGGATTGTGGTGAAC	ATCTGTCCGAGGTGAATCT	148	0.957	0.004	1.170	0.978	1.102	0.504
Coja4716	Up-regulated	TTTAAGGCAACTGACCATCC	CTCTCACTGTTTCTTCCATCC	167	0.987	0.004	1.456	0.504	1.465	0.703
Cyclophilin[c]	Housekeeping	CCCTGACGAGAACTTCAAGC	TTCTCCACCTTCCTCACCAC	200	0.929	0.003				
Glyceraldehyde-3 phosphate dehydrogenase[d]	Housekeeping	GAAGCTTACTGGAATGGCTT	CAGACGGCAGGTCAGGTCA	~100	0.963	0.005				
β-actin	Internal control/ Housekeeping	ATGCAGAAGGAGATCACAGC	TGTTGGTAACAGTCCGGTTT	200	0.952	0.003				

[a] Up-regulated means that the gene was significantly upregulated in SD versus either SD+T or LD males

[b] Average (Avg) and standard deviations (SD) of efficiencies calculated in Real-Time PCRMiner (Zhao & Fernald 2005)

[c] From Foye-Jackson *et al.,* (2011)

[d] From Charlier (2006)

Table S4.2 Percent of males producing foam

	Time points[b]		
Treatment[a]	**1**	**2**	**3**
LD	100%	100%	100%
SD	100%	0%	0%
SD + T	100%	0%	100%

[a] 1 = Baseline; 2 = After photoperiod treatment; 3 = After hormone treatment

[b] LD = long day; SD = short day; SD + T = short day plus testosterone; N = 6 for each treatment

Table S4.3 Annotation of foam proteins

Transcript	*G. gallus* ortholog[a,b]	Protein description[b,c]	emPAI[e]	RPKM[f]	Log FC[h]	ID[i]	SP[j]	d_N[k]	d_S[k]	d_N/d_S[k]	OHR[l]	Re[m]	En[m]
Coja12182_c0_seq1	*00000016761	Lysozyme, g-type 2	975.20	40.84	0.19	P	Y	0.07	0.16	0.48	1.00	U	FG
Coja17575_c0_seq1	No hit	Novel	457.90	14420.80	**3.00**	P, GE	Y					U	FG
Coja10321_c0_seq1	**00000023422*	*Uncharacterized protein*	93.44	645.61	**3.94**	P, GE	N				1.04	U	FG
Coja15808_c0_seq1	*00000016962	Uncharacterized protein	90.95	455.29	**3.77**	P, GE	Y				0.85	U	FG
Coja10892_c0_seq2	*00000013072	Aldo-keto reductase family 1, member B1	62.83	2.67	1.06	P	N	0.02	0.11	0.20	1.00	U	FG
Coja4845_c0_seq1	*00000006453	Ovotransferrin (conalbumin)	56.36	16.60	-0.28	P	Y	0.11	0.23	0.48	1.00	U	M
Coja4828_c0_seq1	*00000001381	Actin, cytoplasmic type 5	54.06	10.23	-0.24	P	N				1.00	U	
Coja21541_c0_seq1	*00000003770	Annexin A2	49.06	8.64	-0.29	P	N	0.00	0.07	0.06	1.00	U	FG
Coja12258_c0_seq1	**00000003506*	*Similar to ovoinhibitor precursor*	48.04	182.02	-0.92	P	N	0.08	0.25	0.33	1.00	U	M
Coja10353_c0_seq1	Not ortholog		38.71	0.99	-0.04	P	N				0.44	FG	FG
Coja18772_c0_seq1	*00000024011	Uncharacterized protein	37.72	21.74	-0.04	P	Y				0.87	U	FG
Coja22380_c0_seq1	*00000015148	Annexin A1	35.86	9.96	0.46	P	N	0.02	0.10	0.18	1.00	U	FG
Coja7727_c0_seq1	Not ortholog		34.79	0.33	0.18	P	N				0.23	U	
Coja17586_c0_seq1	*00000020180	Albumin (serum type)	26.32	1.10	-0.30	P	Y	0.06	0.14	0.44	1.00	U	L
Coja8448_c0_seq1	*00000000620	Fatty acid binding protein 3	20.95	1.50	-0.59	P	N	0.01	0.11	0.10	0.99	U	FG
Coja10336_c0_seq1	*00000000919	Polymeric immunoglobulin receptor	14.95	41.49	**1.40**	P, GE	Y	0.13	0.32	0.40	0.82	U	M
Coja9345_c0_seq1	No hit	Novel	14.20	1.89	**-3.68**	P	N					U	FG
Coja4850_c0_seq1	*00000014442	Glyceraldehyde-3-phosphate dehydrogenase	13.19	6.67	**-2.44**	P	N	0.01	0.16	0.04	1.00	U	
Coja12359_c0_seq1	*00000003053	Peroxiredoxin 6	11.19	0.99	-0.15	P	N	0.01	0.17	0.03	1.00	U	
Coja4514_c0_seq1	Not ortholog		10.02	1.35	**-0.80**	P	N				0.36	U	FG

Coja20297_c0_seq1	No hit	Novel	9.06	1.55	-0.09	P	N					U	L
Coja10265_c0_seq1	*00000002932	Nucleoside diphosphate kinase	9.02	5.81	0.22	P	N	0.02	0.27	0.07	0.99	U	
Coja10694_c0_seq1	Not ortholog		8.49	2.03	0.13	P	N				0.47	U	FG
Coja5078_c0_seq1	Not ortholog		7.81	0.72	-0.21	P	N				0.15	U	FG
Coja6886_c0_seq1	*00000011930	Uncharacterized protein	7.50	4.64	0.48	P	Y				0.99	U	FG
Coja354024_c0_seq1	*00000023229	Histone H2B 8	7.26	0.07	0.54	P	N						
Coja20725_c0_seq1	*00000005393	ADP-ribosylation factor 1	7.17	2.39	**0.84**	P, GE	N	0.00	0.11	0.00	0.99	U	
Coja13765_c0_seq1	No hit	Novel	6.79	7.49	0.29	P	Y					U	M
Coja298958_c0_seq1	Not ortholog		6.57	0.23	0.55	P	N						
Coja10313_c0_seq1	Not ortholog		6.37	42.76	**3.37**	P, GE	N				0.64	U	FG
Coja18367_c0_seq1	*00000005002	Voltage-dependent anion channel 2	6.37	2.08	**-0.69**	P	N	0.01	0.06	0.14	0.95	U	T
Coja18748_c0_seq1	*00000008667	Destrin	6.18	6.08	0.35	P	N	0.00	0.05	0.00	0.99	U	
Coja35839_c0_seq1	Not ortholog		5.84	2.43	**0.81**	P, GE	N				0.48	U	FG
Coja4761_c0_seq1	Not ortholog		5.77	2.53	0.01	P	N				0.24	FG	FG
Coja4201_c0_seq1	Not ortholog		5.77	0.35	0.80	P	N				1.00	M	FG
Coja28359_c0_seq1	No hit	Novel	5.74	3.72	**-3.87**	P	N					U	FG
Coja31532_c0_seq1	*00000005309	Transketolase	5.70	0.83	0.32	P	N	0.01	0.10	0.09	1.00	U	
Coja38960_c1_seq1	*00000002797	6-phosphogluconate dehydrogenase, decarboxylating	5.69	0.94	0.63	P	N	0.01	0.10	0.11	1.00	U	
Coja4892_c0_seq1	*00000001262	Chloride intracellular channel 4	5.49	0.57	0.26	P	N	0.01	0.11	0.07	1.00	U	T
Coja12356_c0_seq1	*00000021139	Ig lambda chain C region	5.42	4.66	**-3.89**	P	Y	0.15	0.21	0.71	1.00	U	FG
Coja8331_c0_seq1	*00000012873	Uncharacterized protein	5.35	0.47	0.15	P	N	0.01	0.13	0.06	1.00	M	FG
Coja17725_c0_seq1	*00000002377	Enolase 1	5.13	2.63	-0.06	P	N	0.03	0.12	0.23	0.99	U	L
Coja348236_c0_seq1	Not ortholog		5.02	0.18	0.46	P	N						
Coja17679_c0_seq1	*00000008586	Tubulin, beta 3	4.92	1.96	-0.15	P	N	0.00	0.29	0.00	0.96	U	T
Coja12823_c0_seq1	*00000010364		4.83	0.21	-0.01	P	N				0.98	U	L
Coja12720_c0_seq1	*00000005930	Plastin 3 (T isoform)	4.79	0.75	0.43	P	N	0.00	0.14	0.02	1.00	U	
Coja21285_c0_seq1	*00000004956	Glucose phosphate isomerase	4.75	0.93	-0.17	P	N	0.01	0.11	0.08	1.00	U	

Coja19770_c0_seq1	**00000018396*	*S1 calcium-binding protein A6*	4.67	9.68	-0.60	P	N				0.98	U	FG
Coja18106_c0_seq1	*00000008125	Ribosomal protein S27a	4.42	6.38	-0.09	P	N	0.00	0.41	0.01	0.87	U	
Coja13894_c0_seq1	*00000000081	Interleukin 4 induced 1	4.34	1.23	0.21	P	Y				0.97	U	FG
Coja29958_c0_seq1	Not ortholog		4.32	0.69	**-1.57**	P	N				0.74	U	U
Coja4863_c0_seq1	*00000001949	Malate dehydrogenase	4.31	1.27	-0.24	P	N	0.01	0.13	0.09	1.00	U	L
Coja15607_c0_seq1	*00000016707	Chloride intracellular channel 5	4.27	0.16	0.20	P	N	0.00	0.06	0.00	1.00	FG	FG
Coja20205_c0_seq1	*00000005622	Uncharacterized protein	4.25	3.57	**1.59**	P, GE	N	0.00	0.14	0.02	0.99	U	
Coja12782_c0_seq1	Not ortholog		4.22	0.96	0.29	P	N				0.92	U	
Coja10032_c0_seq1	Not ortholog		4.20	1.24	0.59	P	N				0.43	U	M
Coja10247_c0_seq1	*00000014526	Triosephosphate isomerase	4.15	1.38	**-0.95**	P	N	0.00	0.11	0.04	1.00	U	
Coja11525_c0_seq1	*00000000479	Peptidyl arginine deiminase, type III	4.05	0.32	0.28	P	N	0.03	0.27	0.11	1.00	M	FG
Coja31420_c0_seq1	**00000024134*		3.90	3.38	-0.07	P	N				0.34	FG	FG
Coja21538_c0_seq1	Not ortholog		3.85	3.18	-0.22	P	N				0.99	U	
Coja9498_c0_seq1	Not ortholog		3.82	0.67	**0.62**	P, GE	N				0.49	U	
Coja4697_c0_seq1	*00000017184	Matrix metalloproteinase 7	3.74	5.36	1.84	P	Y	0.06	0.14	0.40	1.00	U	FG
Coja18989_c0_seq1	*00000010243	Peroxiredoxin-1	3.56	2.23	-0.03	P	N	0.01	0.18	0.08	0.99	U	L
Coja16853_c0_seq1	*00000005763	Villin-like	3.55	0.80	**1.34**	P, GE	N	0.01	0.16	0.06	1.00	U	FG
Coja23918_c0_seq1	*00000005127	Transaldolase 1	3.52	1.81	0.06	P	N	0.01	0.17	0.06	1.00	U	
Coja13005_c0_seq1	*00000011885	Annexin A5	3.42	0.84	-0.21	P	N	0.01	0.14	0.07	1.00	U	L
Coja19425_c0_seq1	*00000001992	Pyruvate kinase, muscle	3.38	1.84	**-1.22**	P	N	0.02	0.18	0.09	1.00	U	T
Coja4901_c0_seq1	*00000007403	Phosphatidylethanolamine binding protein 1	3.38	1.07	-0.29	P	N	0.02	0.22	0.07	0.99	U	
Coja14435_c0_seq1	*00000011888	Transglutaminase 4 (prostate)	3.35	0.48	0.21	P	N	0.04	0.19	0.21	0.91	U	FG
Coja5100_c0_seq1	*00000007936	Phosphoglycerate kinase 1	3.13	1.06	-0.15	P	N	0.02	0.07	0.27	1.00	U	
Coja11212_c0_seq1	No hit	Novel	3.12	0.69	**1.08**	P, GE	N					U	FG
Coja37151_c0_seq1	Not ortholog		3.12	0.69	0.38	P	N				0.33	U	
Coja18377_c0_seq1	*00000006512	Heat shock cognate 7	3.06	7.42	0.75	P	N	0.00	0.10	0.00	0.94	U	U

Coja21510_c0_seq1	*00000002661	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, epsilon polypeptide	3.05	2.22	0.36	P	N	0.06	0.08	0.71	0.97	U	
Coja10005_c0_seq1	*00000016535	Ras-related GTP-binding protein RAB1	2.95	1.18	0.37	P	N	0.00	0.11	0.04	0.99	U	U
Coja19610_c0_seq1	*00000008783	GTP binding protein Rab1a	2.92	4.12	**1.30**	P, GE	N	0.00	0.03	0.00	1.00	U	
Coja12882_c0_seq1	*00000014412	Cystatin A (stefin A)	2.87	1.09	0.34	P	N	0.13	0.14	0.93	0.97	U	FG
Coja5133_c0_seq1	*00000009070	Heat shock 1kDa protein 1 (chaperonin 1)	2.87	1.01	-0.30	P	N	0.00	0.13	0.00	0.99	U	L
Coja23431_c0_seq1	*00000010725	14-3-3 protein zeta	2.86	1.82	0.38	P	N	0.00	0.05	0.00	0.95	U	
Coja10343_c0_seq1	Not ortholog		2.79	2.74	0.02	P	N				0.61	U	
Coja32892_c0_seq1	*00000016153	Prostate stem cell antigen	2.74	2.85	-0.63	P	Y	0.09	0.08	1.08	0.99	M	FG
Coja10911_c0_seq2	Not ortholog		2.69	1.07	-0.09	P	N				0.19	U	FG
Coja13044_c0_seq1	*00000009474	Alpha-actinin-1	2.69	0.80	0.24	P	N	0.00	0.07	0.01	1.00	U	
Coja21368_c0_seq1	*00000015147	Aldehyde dehydrogenase 1A1	2.69	0.33	-0.78	P	N	0.07	0.20	0.36	1.00	U	L
Coja15798_c0_seq1	Not ortholog		2.60	0.63	0.10	P	N				0.97	U	FG
Coja7446_c0_seq1	Not ortholog		2.47	0.60	0.02	P	N				0.52	U	
Coja22163_c0_seq1	*00000004143	14-3-3 protein beta/alpha	2.45	0.76	0.33	P	N	0.00	0.07	0.00	1.00	U	
Coja5916_c0_seq1	*00000016433	14-3-3 protein theta	2.43	1.74	0.43	P	N	0.00	0.05	0.00	0.93	U	
Coja21166_c0_seq1	*00000008818	Isocitrate dehydrogenase	2.40	0.41	0.49	P	N	0.00	0.10	0.02	0.99	U	L
Coja15661_c0_seq1	*00000019795	Uncharacterized protein	2.39	0.46	-0.10	P	N				1.10	U	M
Coja4927_c0_seq1	Not ortholog		2.38	3.36	0.23	P	N				0.99	U	
Coja22470_c0_seq1	*00000006300	Lactate dehydrogenase A	2.36	2.06	-0.20	P	N	0.01	0.10	0.05	1.00	U	FG
Coja9672_c0_seq1	Not ortholog		2.36	0.44	0.67	P	N				1.00	U	FG
Coja17645_c0_seq1	*00000015917	Elongation factor 1-alpha 1	2.33	25.41	-0.08	P	N	0.00	0.10	0.01	1.00	U	
Coja18197_c0_seq1	*00000007606	Phosphoglycerate mutase 1	2.31	1.55	-0.24	P	N	0.01	0.30	0.05	0.99	U	

Coja18295_c0_seq1	*00000010439	PREDICTED: hypothetical protein	2.31	0.03	-0.42	P	N	0.00	0.16	0.00	0.98	M	T
Coja17681_c0_seq1	*00000007468	Alpha 1 globin	2.30	3.69	-0.35	P	N	0.04	0.12	0.30	0.98	U	L
Coja12659_c0_seq1	*00000008169	GDP dissociation inhibitor 2	2.27	1.40	**0.82**	P, GE	N	0.02	0.13	0.12	1.00	U	
Coja9182_c0_seq1	Not ortholog		2.21	1.57	0.13	P	N				0.05	FG	FG
Coja31501_c0_seq1	Not ortholog		2.20	1.73	-0.57	P	N				0.46	U	FG
Coja584482_c0_seq1	Not ortholog		2.16	0.02	0.20	P	N						
Coja18559_c0_seq1	No hit	Novel	2.15	4.20	-0.08	P	N					U	
Coja4765_c0_seq1	*00000002859	Ras-related C3 botulinum toxin substrate 3 (rho family, small GTP binding protein Rac3)	2.13	1.79	0.32	P	N				0.85	U	
Coja10342_c0_seq1	*00000022539	ATP synthase subunit beta, mitochondrial	2.10	2.66	-0.34	P	N				1.00	U	L
Coja14772_c0_seq1	*00000005956	Uncharacterized protein	2.05	1.19	1.17	P	N	0.02	0.10	0.18	1.00	FG	FG
Coja17000_c0_seq1	*00000010481	UMP-CMP kinase	2.05	0.59	**1.44**	P, GE	N	0.00	0.07	0.06	0.99	U	
Coja14629_c0_seq1	*00000012866	Serine proteinase inhibitor, clade B (ovalbumin), member 6 (SERPIN B6)	2.01	0.55	0.74	P	N	0.03	0.14	0.24	1.00	U	
Coja4889_c0_seq1	*00000003589	Vitronectin	2.00	1.96	-1.08	P	Y	0.04	0.23	0.17	0.96	U	L
Coja8317_c0_seq1	*00000019430	Uncharacterized protein	2.00	0.49	0.09	P	Y				0.99	M	FG
Coja5202_c0_seq1	Not ortholog		1.91	4.13	-0.10	P	N				0.96	U	U
Coja5099_c0_seq1	*00000008341	Furin (paired basic amino acid cleaving enzyme)	1.89	3.95	**3.68**	P, GE	Y	0.01	0.20	0.04	1.00	U	FG
Coja4866_c0_seq1	*00000004725	Uncharacterized protein	1.89	1.45	0.11	P	N	0.01	0.12	0.09	1.00	U	L
Coja10319_c0_seq1	Not ortholog		1.87	7.32	0.34	P	N				1.00	U	L
Coja36090_c0_seq1	Not ortholog		1.87	1.06	0.06	P	N				0.14	U	FG

Coja21594_c0_seq1	*00000010614	Solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 4	1.86	1.06	**1.20**	P, GE	N				1.00	U	
Coja18020_c0_seq1	*00000013990	4S ribosomal protein S12	1.80	12.63	-0.25	P	N	0.00	0.15	0.02	0.99	U	
Coja28237_c0_seq1	*00000002351	Purine nucleoside phosphorylase	1.79	0.23	**1.52**	P, GE	N				0.86	U	L
Coja18172_c0_seq1	*00000006771	PREDICTED: similar to ribosomal protein S15 isoform 1	1.78	7.62	-0.55	P	N	0.00	0.07	0.00	0.95	U	
Coja23138_c0_seq1	*00000004034	F-actin-capping protein subunit beta isoforms 1 and 2	1.78	1.15	-0.10	P	N	0.00	0.02	0.00	1.00	U	
Coja15510_c0_seq1	*00000013742	Ezrin	1.77	1.23	**1.34**	P, GE	N	0.00	0.14	0.03	1.00	U	FG
Coja12158_c0_seq1	*00000007233	Protein disulfide-isomerase	1.71	6.55	**2.11**	P, GE	Y	0.01	0.13	0.05	1.00	U	
Coja16284_c0_seq1	*00000004435	Rab27aUncharacterized protein	1.70	1.13	**1.05**	P, GE	N	0.00	0.10	0.00	1.00	U	FG
Coja4797_c0_seq1	*00000003806	Ras homolog gene family, member A	1.69	4.89	0.54	P	N	0.00	0.08	0.00	0.94	U	
Coja18805_c0_seq1	*00000004449	Cytochrome P45 A 37	1.68	0.97	0.59	P	N	0.05	0.16	0.31	1.00	U	L
Coja19688_c0_seq1	*00000019445	Cytochrome c oxidase subunit VIc	1.65	1.88	-0.43	P	N	0.01	0.10	0.07	1.00	U	L
Coja22982_c0_seq1	*00000016357	Peroxiredoxin 4	1.65	1.73	**1.39**	P, GE	Y	0.01	0.15	0.08	0.90	U	L
Coja19840_c0_seq1	*00000013098	Microsomal glutathione S-transferase 1	1.64	1.76	**1.04**	P, GE	N	0.02	0.10	0.16	0.99	U	L
Coja19281_c0_seq1	*00000005338	Ribosomal protein S21	1.62	2.46	-0.60	P	N	0.04	0.27	0.14	0.94	U	
Coja18619_c0_seq1	*00000015844	Superoxide dismutase 1, soluble	1.62	0.92	-0.44	P	N	0.00	0.14	0.00	0.99	U	L

Coja12120_c0_seq1	*00000012173	Lectin, galactoside-binding, soluble, 3 (galectin 3)	1.62	0.68	0.05	P	N	0.06	0.36	0.17	0.69	U	FG
Coja38482_c0_seq1	*00000015702	NADP-dependent leukotriene B4 12-hydroxydehydrogenase	1.62	0.20	0.26	P	N	0.03	0.15	0.22	1.00	U	L
Coja17522_c0_seq1	*00000012790	Desmoplakin	1.58	0.69	0.67	P	N	0.01	0.10	0.09	0.96	U	FG
Coja4777_c0_seq1	*00000006691		1.57	1.02	-0.28	P	N				0.99	U	
Coja4840_c0_seq1	*00000017330	Ribosomal protein S3	1.55	6.65	-0.21	P	N	0.00	0.09	0.00	1.00	U	
Coja14110_c0_seq1	No hit	Novel	1.54	0.49	0.67	P	N					U	L
Coja15308_c0_seq1	*00000007406	Uncharacterized protein	1.53	0.50	0.07	P	N	0.04	0.31	0.14	0.93	FG	FG
Coja10478_c0_seq1	Not ortholog		1.53	0.08	-0.17	P	N				0.31	L	L
Coja4373_c0_seq1	*00000004571	Protein phosphatase 1, catalytic subunit, gamma isoform	1.52	0.47	-0.01	P	N	0.00	0.14	0.00	1.00	U	
Coja13713_c0_seq1	*00000001434	Stomatin	1.50	0.29	-0.94	P	N	0.03	0.13	0.24	1.00	U	
Coja61441_c0_seq1	*00000023622	Avidin	1.49	0.64	0.87	P	Y	0.07	0.12	0.59	0.83	M	FG
Coja4823_c0_seq1	*00000001830	Eukaryotic translation elongation factor 2	1.47	5.68	0.00	P	N	0.01	0.24	0.02	1.00	U	U
Coja5295_c0_seq1	*00000003820	Actin related protein 2/3 complex, subunit 3, 21kDa	1.46	1.01	0.30	P	N	0.00	0.12	0.03	0.99	U	
Coja18368_c0_seq1	*00000006096	4S ribosomal protein S13	1.45	6.18	-0.39	P	N	0.00	0.12	0.00	0.99	U	
Coja12443_c0_seq1	*00000001000	78 kDa glucose-regulated protein	1.44	7.01	**1.78**	P, GE	Y	0.00	0.18	0.00	1.00	U	
Coja5130_c0_seq2	Not ortholog		1.41	1.97	**1.26**	P, GE	N				0.88	U	FG
Coja10516_c0_seq1	Not ortholog		1.38	0.66	0.13	P	Y						
Coja17687_c0_seq1	*00000013257	L-lactate dehydrogenase B chain	1.37	0.71	**-1.82**	P	N	0.04	0.12	0.36	1.00	U	T
Coja18375_c0_seq1	*00000008806	Uncharacterized protein	1.36	6.80	-0.43	P	N	0.02	0.26	0.06	0.98	U	

Coja31572_c0_seq1	Not ortholog		1.36	0.68	**0.49**	P, GE	N				0.59	U	
Coja68414_c0_seq1	*00000001231	SAM domain- and HD domain-containing protein 1	1.36	0.35	0.21	P	N	0.02	0.19	0.13	0.83	U	FG
Coja16521_c0_seq1	*00000019802	Pleckstrin homology-like domain, family A, member 3	1.36	0.06	0.22	P	N				1.17		FG
Coja18507_c0_seq1	*00000004831	4S ribosomal protein S4	1.33	6.87	-0.35	P	N	0.00	0.16	0.00	1.00	U	
Coja13097_c0_seq1	*00000015079	Uroplakin 1B	1.33	1.38	0.89	P	N	0.19	0.34	0.56	1.00	FG	FG
Coja21084_c0_seq1	*00000001986	Valosin-containing protein	1.28	1.02	0.37	P	N	0.01	0.13	0.04	1.00	U	
Coja18728_c0_seq1	*00000008684	Eukaryotic initiation factor 4A-II	1.27	8.38	0.79	P	N	0.00	0.08	0.00	1.00	U	
Coja204259_c0_seq1	*00000013073	Uridine phosphorylase 1	1.24	0.09	0.03	P	N	0.01	0.07	0.11	0.90	U	
Coja4716_c0_seq1	*00000011806	Mucin 13, cell surface associated	1.23	4.51	**1.46**	P, GE	N	0.18	0.23	0.79	0.62	M	FG
Coja18070_c0_seq1	**00000014644*	*ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit*	1.23	3.96	-0.19	P	N	0.00	0.11	0.01	1.00	U	
Coja10745_c0_seq1	*00000000613	RAB11B, member RAS oncogene family	1.23	0.33	-0.42	P	N	0.00	0.20	0.00	1.00	U	
Coja13458_c0_seq1	Not ortholog		1.23	0.23	0.37	P	N				0.79	M	M
Coja13883_c0_seq1	Not ortholog		1.21	1.00	**-3.27**	P	N				1.00	U	
Coja11350_c0_seq1	Not ortholog		1.17	0.90	1.01	P	N				0.06	M	FG
Coja12992_c0_seq1	*00000013033	Uncharacterized protein	1.17	0.44	0.64	P	N				1.00	M	M
Coja11151_c0_seq1	Not ortholog		1.17	0.38	-0.46	P	N				1.00	U	U
Coja32501_c1_seq1	Not ortholog		1.16	0.80	0.42	P	N				0.98	U	
Coja19098_c0_seq1	*00000002569	GTP-binding nuclear protein Ran	1.15	1.12	-0.02	P	N	0.00	0.08	0.00	0.99	U	T
Coja10396_c0_seq1	*00000011020	Cytochrome c	1.14	1.60	0.20	P	N	0.01	0.11	0.08	0.97	U	
Coja5178_c0_seq1	*00000014567	Prohibitin 2	1.14	0.68	-0.27	P	N	0.01	0.17	0.05	1.00	U	

Coja147948_c0_seq1	Not ortholog		1.13	0.26	**1.82**	P, GE	N				0.43	U	FG
Coja4834_c0_seq1	*00000017172	Vitelline membrane outer layer protein 1	1.13	0.04	-0.75	P	Y				0.92	U	L
Coja4818_c0_seq1	*00000011612	Group-specific component (vitamin D binding protein)	1.12	0.05	-0.43	P	Y	0.04	0.17	0.23	0.96	U	L
Coja5145_c0_seq1	*00000010332*	*Uncharacterized protein*	1.11	5.58	-0.53	P	N	0.00	0.08	0.00	0.84	U	
Coja14721_c0_seq1	*00000001446	Gelsolin	1.11	2.08	**-1.26**	P	N	0.01	0.14	0.06	1.00	U	FG
Coja12493_c0_seq1	*00000013400	Programmed cell death 6	1.11	1.06	0.47	P	N	0.01	0.22	0.03	0.95	U	
Coja10220_c0_seq1	*00000006022	Proteasome (prosome, macropain) subunit, alpha type, 1	1.11	0.79	0.05	P	N	0.00	0.13	0.00	1.00	U	
Coja15902_c0_seq1	*00000007152	Family with sequence similarity 3, member D	1.11	0.57	0.38	P	N	0.05	0.17	0.27	1.00	M	FG
Coja279889_c0_seq1	Not ortholog		1.09	0.00		P	N				0.26	T	T
Coja10022_c0_seq1	*00000008969	Tumor-associated calcium signal transducer 1	1.07	4.96	0.51	P	Y	0.09	0.17	0.53	1.00	U	FG
Coja20366_c0_seq1	*00000016027	Carbonyl reductase 1	1.06	0.74	**1.03**	P, GE	N	0.02	0.15	0.15	0.99	U	L
Coja4720_c0_seq1	*00000006326	Macrophage migration inhibitory factor	1.05	4.05	**1.02**	P, GE	N	0.00	0.39	0.00	0.99	U	
Coja17592_c0_seq1	*00000007114	Apolipoprotein A-I	1.05	3.09	**-2.02**	P	Y	0.04	0.26	0.14	1.00	U	L
Coja9990_c0_seq1	*00000022421*		1.05	0.76	**-2.88**	P	Y				0.63	U	FG
Coja17398_c0_seq1	*00000019514	CNDP dipeptidase 2 (metallopeptidase M2 family)	1.05	0.68	**1.86**	P, GE	N	0.02	0.12	0.21	0.99	U	

Coja16436_c0_seq1	*00000011292	Ubiquitin-conjugating enzyme E2N (UBC13 homolog, yeast)	1.05	0.48	0.13	P	N	0.03	0.14	0.25	0.96	U	
Coja7632_c0_seq1	*00000008087	Carboxypeptidase A1 (pancreatic)	1.05	0.00		P	N				1.00	T	T
Coja15108_c0_seq1	Not ortholog		1.04	0.82	0.20	P	N				1.00	U	U
Coja17103_c0_seq1	*00000007128	SH3 domain binding glutamic acid-rich protein like	1.04	0.34	0.04	P	N	0.08	0.36	0.22	0.94	U	FG
Coja10971_c0_seq1	**00000019162*	*GM2 ganglioside activator*	1.03	0.12	-0.47	P	Y	0.04	0.41	0.11	0.98	U	L
Coja28027_c0_seq1	Not ortholog		1.02	0.93	**-1.63**	P	N				0.18	U	
Coja394045_c0_seq1	Not ortholog		1.02	0.15	0.64	P	N						
Coja408087_c0_seq1	No hit	Novel	1.02	0.08	-0.57	P	N						
Coja18625_c0_seq1	*00000008858	Malate dehydrogenase, cytoplasmic	1.01	1.15	0.14	P	N	0.00	0.10	0.05	1.00	U	
Coja11210_c0_seq1	*00000008382	Guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 1	1.01	0.29	**1.27**	P, GE	N	0.00	0.10	0.00	1.00	M	FG
Coja9037_c0_seq1	*00000005346	Uncharacterized protein	0.99	1.21	**1.09**	P, GE	Y	0.04	0.31	0.13	0.93	U	FG
Coja72091_c0_seq1	Not ortholog		0.99	0.36	-1.09	P	N				0.21	U	
Coja24402_c0_seq1	Not ortholog		0.98	0.88	0.00	P	N				0.20	U	U
Coja18222_c0_seq1	*00000014432	Ribosomal protein S2	0.97	10.41	-0.49	P	N	0.00	0.12	0.00	0.99	U	
Coja18354_c0_seq1	*00000008348	Protein disulfide-isomerase A3	0.97	5.47	**1.28**	P, GE	Y	0.01	0.13	0.11	0.95	U	
Coja11040_c0_seq1	**00000007280*	*Claudin 1*	0.97	1.24	0.39	P	N	0.03	0.25	0.11	1.00	U	FG
Coja33422_c0_seq1	*00000007435	Ras-related protein Rab-18	0.96	0.53	**0.53**	P, GE	N	0.00	0.05	0.00	1.00	U	
Coja68250_c0_seq1	Not ortholog		0.96	0.36	0.37	P	N				0.24	U	

Coja6027_c0_seq1	*00000007615	Ras-related protein Rab-11A	0.95	0.94	**0.67**	P, GE	N	0.00	0.04	0.00	1.00	U	
Coja5666_c0_seq1	*00000007308	RAB35, member RAS oncogene family	0.95	0.17	0.41	P	N	0.01	0.06	0.11	1.00	U	
Coja12878_c0_seq1	Not ortholog		0.94	0.50	-0.53	P	Y				0.79	M	FG
Coja8231_c0_seq1	*00000006743	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, eta polypeptide	0.94	0.36	-0.02	P	N	0.00	0.11	0.00	1.00	U	
Coja10567_c0_seq1	No hit	Novel	0.93	2.07	**1.55**	P, GE	Y					FG	FG
Coja17513_c0_seq1	*00000023872	Family with sequence similarity 129, member B	0.92	0.81	0.32	P	N	0.06	0.18	0.34	0.59	U	FG
Coja11033_c0_seq1	*00000001474	Ras-related protein Rab-14	0.92	0.68	**0.48**	P, GE	N	0.00	0.07	0.00	1.00	U	
Coja23422_c0_seq1	**00000023484*	*Thioredoxin domain containing 17*	0.89	1.65	**0.68**	P, GE	N	0.01	0.08	0.12	1.00	U	L
Coja12631_c0_seq1	*00000008779	Actin-related protein 2	0.89	0.80	**0.89**	P, GE	N	0.02	0.13	0.17	0.96	U	
Coja27762_c0_seq1	*00000015450	Ras-related protein Rab-2A	0.89	0.70	0.40	P	N	0.00	0.06	0.00	0.99	U	
Coja11053_c0_seq1	Not ortholog		0.89	0.47	-0.53	P	N				0.38	U	
Coja27155_c0_seq1	*00000014320	UDP-glucose 6-dehydrogenase	0.88	1.33	**1.49**	P, GE	N	0.00	0.15	0.01	1.00	U	FG
Coja15615_c0_seq1	*00000012043	Programmed cell death 6 interacting protein	0.88	0.57	**0.89**	P, GE	N	0.00	0.14	0.03	0.85	U	
Coja1137239_c0_seq1	No hit	Novel	0.88	0.01		P	N						
Coja5788_c0_seq1	*00000015320	N-acetylneuraminic acid synthase (sialic acid synthase)	0.87	2.14	**1.30**	P, GE	N	0.04	0.24	0.17	1.00	U	FG

Coja8701_c0_seq1	*00000006520	Myosin-11	0.87	0.99	0.55	P	N				1.00	U	
Coja12648_c0_seq1	Not ortholog		0.87	0.91	0.21	P	N				0.70	U	
Coja12815_c0_seq1	*00000007650	Solute carrier family 9 (sodium/hydrogen exchanger), isoform 3 regulator 1	0.87	0.23	-0.26	P	N	0.03	0.42	0.07	1.00	U	L
Coja12004_c1_seq1	Not ortholog		0.86	0.59	0.21	P	N				1.00	U	
Coja14614_c0_seq1	*00000002201	Uncharacterized protein	0.84	1.18	0.17	P	N	0.02	0.17	0.11	1.01	M	FG
Coja14855_c0_seq1	*00000007129	Aldehyde dehydrogenase 1 family, member A3	0.84	0.77	**0.89**	P, GE	N	0.00	0.12	0.01	0.82	M	FG
Coja11521_c0_seq1	*00000011792	Uncharacterized protein	0.84	0.48	0.43	P	N	0.04	0.13	0.29	0.98	M	FG
Coja5720_c0_seq1	*00000004693	Actin-related protein 2/3 complex subunit 1A	0.83	1.23	**0.50**	P, GE	N	0.00	0.11	0.00	1.00	U	
Coja4896_c0_seq1	*00000003493*	*Ribosomal protein S27*	0.83	0.16	**-0.97**	P	N				1.00	U	L
Coja51151_c0_seq1	*00000007488	Zinc finger protein 185 (LIM domain)	0.81	0.64	**1.37**	P, GE	N				0.53	M	FG
Coja23060_c0_seq1	*00000001476	RAP1A, member of RAS oncogene family	0.80	0.89	**0.51**	P, GE	N	0.00	0.07	0.03	0.97	U	
Coja4871_c0_seq1	*00000002490	Uncharacterized protein	0.80	0.75	-0.50	P	N	0.05	0.34	0.16	1.00	U	
Coja62843_c0_seq1	Not ortholog		0.79	0.41	0.08	P	N				0.33	U	FG
Coja382690_c0_seq1	No hit	Novel	0.79	0.09	1.01	P	N						
Coja29395_c0_seq1	*00000009385	F-actin-capping protein subunit alpha-2	0.78	0.81	0.47	P	N	0.01	0.23	0.04	1.00	U	
Coja13777_c0_seq1	*00000010770	Adseverin	0.78	0.26	**-0.93**	P	N	0.01	0.10	0.10	0.91	U	FG
Coja10646_c0_seq1	*00000006947	3-hydroxybutyrate dehydrogenase	0.77	1.21	**1.62**	P, GE	N	0.02	0.17	0.14	1.01	U	FG

Coja5200_c0_seq1	*00000007484	Glutamic-oxaloacetic transaminase 1, soluble (aspartate aminotransferase 1)	0.77	0.24	**-0.85**	P	N	0.01	0.11	0.10	0.99	U	
Coja4794_c0_seq1	*00000002317	Unnamed protein product	0.77	0.04	**-1.12**	P	N	0.02	0.13	0.18	0.95	U	L
Coja18139_c0_seq1	*00000016392	Ribosomal protein S7	0.76	6.16	-0.39	P	N	0.00	0.03	0.00	0.96	U	
Coja17651_c0_seq1	*00000015143	Transthyretin	0.76	0.23	-0.38	P	Y	0.00	0.07	0.05	0.99	U	L
Coja10168_c0_seq1	Not ortholog		0.76	0.06	0.02	P	N				0.55	U	L
Coja25984_c0_seq1	No hit	Novel	0.75	0.49	0.02	P	Y					U	
Coja83468_c0_seq1	*00000016471	Uncharacterized protein	0.75	0.33	-0.20	P	N				0.99	U	FG
Coja27058_c0_seq1	**00000005532*	*Calcium binding protein 39*	0.74	0.66	**1.16**	P, GE	N	0.06	0.28	0.21	1.00	U	
Coja49548_c0_seq1	Not ortholog		0.74	0.59	0.10	P	N				0.70	U	FG
Coja663786_c0_seq1	Not ortholog		0.74	0.00		P	N						
Coja23832_c0_seq1	*00000011452	Actin related protein 2/3 complex, subunit 2, 34kDa	0.73	1.45	**0.80**	P, GE	N	0.00	0.13	0.02	0.97	U	
Coja28009_c0_seq1	*00000013723	2'-5'-oligoadenylate synthetase-like	0.73	0.49	0.04	P	N	0.09	0.37	0.25	0.97	U	M
Coja10479_c0_seq1	Not ortholog		0.72	1.86	-0.10	P	N				0.97	U	
Coja13490_c0_seq1	*00000023760	Eosinophil chemotactic cytokine	0.72	1.30	**2.34**	P, GE	Y	0.01	0.16	0.09	1.00	U	FG
Coja10164_c0_seq1	**00000000866*		0.72	1.28	**1.35**	P, GE	N				0.99	U	
Coja11098_c0_seq1	*00000014410	Cystatin B (stefin B)	0.72	1.11	0.39	P	N	0.03	0.10	0.27	0.96	U	
Coja17230_c0_seq1	*00000013229	Cytidine monophosphate N-acetylneuraminic acid synthetase	0.72	0.67	**1.16**	P, GE	N	0.01	0.13	0.06	1.00	U	FG
Coja15838_c0_seq1	Not ortholog		0.72	0.57	0.53	P	N				0.25	M	FG
Coja12954_c0_seq1	Not ortholog		0.72	0.50	**0.67**	P, GE	N				0.70	U	FG
Coja16770_c0_seq1	*00000005349	Mesothelin	0.72	0.35	0.27	P	Y	0.06	0.12	0.45	0.96	M	FG

Coja956346_c0_seq1	No hit	Novel	0.72	0.00		P	N						
Coja10254_c0_seq1	*00000000122	Guanine nucleotide-binding protein subunit beta-2-like 1	0.71	5.85	-0.34	P	N	0.00	0.19	0.00	0.99	U	
Coja5029_c0_seq1	Not ortholog		0.71	3.33	0.01	P	N				0.75	U	
Coja10174_c0_seq1	*00000008443	Rho GDP dissociation inhibitor (GDI) alpha	0.71	1.35	0.43	P	N	0.00	0.12	0.00	1.00	U	
Coja23897_c0_seq1	*00000015862	Carbonic anhydrase 2	0.71	1.19	**1.97**	P, GE	N	0.02	0.07	0.23	1.00	U	
Coja4810_c0_seq1	Not ortholog		0.71	0.67	-0.22	P	N				1.02	U	T
Coja26800_c0_seq1	*00000017135	Lambda-crystallin	0.71	0.38	**1.41**	P, GE	N	0.02	0.09	0.25	0.98	U	L
Coja13544_c0_seq1	*00000002745	Golgi apparatus protein 1	0.70	0.42	**0.83**	P, GE	N	0.02	0.27	0.08	0.93	U	
Coja24670_c0_seq1	*00000005902	Uncharacterized protein	0.69	1.29	0.41	P	N	0.00	0.08	0.00	1.00	U	
Coja5072_c0_seq1	*00000010273	Serine/threonine-protein phosphatase 2A catalytic subunit alpha isoform	0.69	0.76	**0.68**	P, GE	N	0.00	0.13	0.00	1.00	U	
Coja142551_c0_seq1	*00000016502	Tumor protein p53 inducible protein 3	0.69	0.28	**1.48**	P, GE	N	0.01	0.07	0.17	0.99	M	FG
Coja16669_c0_seq1	*00000009065	Hypothetical protein LOC422385	0.69	0.25	**0.80**	P, GE	N	0.01	0.10	0.15	0.99	U	
Coja69089_c0_seq1	*00000004709	Hypothetical protein LOC422173	0.68	0.15	-0.36	P	N	0.00	0.10	0.03	1.00	U	U
Coja20443_c0_seq1	*00000006503	Voltage-dependent anion channel 1	0.67	0.66	-0.40	P	N	0.00	0.08	0.06	1.00	U	L
Coja17388_c0_seq1	*00000002709	Myosin IC	0.67	0.26	0.05	P	N	0.00	0.15	0.02	0.99	U	
Coja113755_c0_seq1	*00000012300	Uncharacterized protein	0.67	0.09	0.40	P	N	0.00	0.19	0.02	0.96	U	
Coja7136_c0_seq1	Not ortholog		0.66	0.25	-0.16	P	N				0.11	U	
Coja264085_c0_seq1	Not ortholog		0.66	0.03	0.20	P	N				0.58	U	U

Coja14781_c0_seq1	*00000010318	Transmembrane emp24 domain-containing protein 1	0.65	3.58	**1.04**	P, GE	Y	0.01	0.11	0.05	0.87	U	
Coja11870_c0_seq1	*00000016555	Uncharacterized protein	0.65	0.42	**-1.18**	P	N	0.01	0.15	0.07	1.00	U	L
Coja5408_c0_seq1	*00000002053	NADPH--cytochrome P45 reductase	0.65	0.33	**0.74**	P, GE	N	0.01	0.14	0.08	0.99	U	L
Coja16033_c0_seq1	*00000004180	Torsin family 1, member B (torsin B)	0.65	0.24	0.56	P	Y	0.02	0.42	0.04	0.91	U	
Coja6676_c0_seq1	*00000017274	CD9 antigen	0.64	1.82	0.04	P	N	0.04	0.08	0.50	1.00	U	FG
Coja20898_c0_seq1	*00000000496*	*Prostaglandin E synthase 3*	0.64	1.76	0.24	P	N	0.00	0.09	0.00	0.99	U	
Coja9556_c0_seq1	Not ortholog		0.64	0.69	-0.10	P	N				0.25	U	FG
Coja11981_c0_seq1	Not ortholog		0.64	0.24	-0.16	P	N				0.39	U	
Coja18016_c0_seq1	*00000007711	Ribosomal protein L4	0.63	7.60	-0.17	P	N	0.01	0.24	0.05	0.82	U	
Coja10346_c0_seq1	*00000004769	Prosaposin	0.63	4.63	0.71	P	Y	0.01	0.13	0.11	0.97	U	
Coja4384_c0_seq1	Not ortholog		0.63	2.17	0.11	P	N				0.81	U	FG
Coja10709_c0_seq1	*00000007493	Uncharacterized protein	0.63	0.13	0.20	P	N	0.02	0.13	0.15	1.00	U	L
Coja18986_c0_seq1	*00000003445	Microsomal glutathione S-transferase 3	0.62	0.70	-0.33	P	N	0.05	0.18	0.28	0.99	U	T
Coja17876_c0_seq1	Not ortholog		0.61	14.83	**3.43**	P, GE	N				0.46	U	FG
Coja18519_c0_seq1	*00000010979	Uncharacterized protein	0.61	9.37	**2.31**	P, GE	N	0.04	0.15	0.30	1.00	U	M
Coja18560_c0_seq1	*00000005749	Similar to cytochrome c oxidase subunit IV precursor	0.61	2.58	**-0.81**	P	N	0.02	0.17	0.09	0.99	U	L
Coja4796_c0_seq1	*00000011514	Solute carrier family 25 (mitochondrial carrier; phosphate carrier), member 3	0.61	2.57	0.05	P	N	0.02	0.23	0.11	1.00	U	
Coja4387_c0_seq1	*00000010825	Anterior gradient 2 homolog (Xenopus laevis)	0.61	2.17	0.27	P	Y	0.01	0.12	0.07	0.99	U	FG

Coja56523_c0_seq1	*00000004243	Uncharacterized protein	0.61	0.26	0.38	P	N				0.85	M	FG
Coja654155_c0_seq1	*00000016969	Laccase (multicopper oxidoreductase) domain containing 1	0.61	0.03	0.61	P	N						
Coja18383_c0_seq1	*00000008212	6S ribosomal protein L3	0.60	7.31	-0.53	P	N	0.00	0.15	0.00	0.99	U	
Coja24342_c0_seq1	*00000010073	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase	0.60	1.87	**1.00**	P, GE	N	0.01	0.25	0.06	0.91	U	
Coja10029_c0_seq1	*00000002321	Glutamic-oxaloacetic transaminase 2, mitochondrial (aspartate aminotransferase 2)	0.60	0.50	-0.40	P	N	0.01	0.13	0.05	0.99	U	L
Coja10374_c0_seq1	*00000005922	6S ribosomal protein L5	0.59	7.22	-0.54	P	N	0.00	0.17	0.02	1.00	U	
Coja15815_c0_seq1	Not ortholog		0.59	5.03	**3.56**	P, GE	Y				0.35	U	FG
Coja4856_c0_seq1	Not ortholog		0.59	3.39	0.38	P	N				0.95	U	
Coja5811_c0_seq1	*00000012159	Actin-related protein 3	0.59	0.72	0.52	P	N	0.00	0.07	0.00	1.00	U	
Coja5143_c0_seq1	**00000000537*	*Parkinson disease (autosomal recessive, early onset) 7*	0.59	0.71	-0.37	P	N	0.01	0.16	0.07	0.99	U	L
Coja15341_c0_seq1	*00000004222	Hexokinase 1	0.59	0.58	**-1.00**	P	N	0.02	0.15	0.15	0.99	U	FG
Coja3362_c0_seq1	*00000016920	LIM domain 7	0.59	0.50	0.34	P	N	0.02	0.09	0.24	0.42	U	FG
Coja10143_c0_seq1	No hit	Novel	0.59	0.43	-0.15	P	N					U	L
Coja18714_c0_seq1	No hit	Novel	0.58	4.63	-0.46	P	N					U	
Coja768631_c0_seq1	No hit	Novel	0.58	0.00		P	N						
Coja10408_c0_seq1	**00000024057*	*Mesothelin-like*	0.57	4.29	0.52	P	Y	0.06	0.14	0.43	1.00	M	FG
Coja5242_c0_seq1	Not ortholog		0.57	2.10	0.30	P	N				0.14	FG	FG
Coja20534_c0_seq1	*00000014023	Histone H2A.Z	0.57	0.80	0.51	P	N	0.00	0.10	0.00	0.99	U	
Coja4704_c0_seq1	*00000010060	Proteasome subunit alpha type	0.57	0.77	-0.08	P	N	0.00	0.07	0.00	0.98	U	
Coja12732_c0_seq1	Not ortholog		0.57	0.70	-0.27	P	N				0.80	U	U

Coja13003_c1_seq1	*00000014142	Uncharacterized protein	0.57	0.59	-0.06	P	N	0.01	0.07	0.14	0.93	U	U
Coja16997_c0_seq1	Not ortholog		0.57	0.49	**1.01**	P, GE	N				0.92	U	FG
Coja8383_c0_seq1	Not ortholog		0.57	0.27	0.01	P	N				0.78	U	FG
Coja10146_c0_seq1	No hit	Novel	0.56	0.84	0.06	P	N					U	
Coja20847_c0_seq1	*00000013045	Tubulin, alpha 8	0.56	0.02	-0.46	P	N	0.01	0.16	0.05	1.00	U	T
Coja5848_c0_seq1	*00000002720	Aquaporin 5	0.55	3.85	**1.14**	P, GE	N	0.03	0.19	0.16	0.91	FG	FG
Coja16152_c0_seq1	*00000006613	Cathepsin DCathepsin D light chainCathepsin D heavy chain	0.55	0.48	0.15	P	Y	0.01	0.17	0.04	0.94	U	L
Coja9197_c0_seq1	*00000006596	Small nuclear ribonucleoprotein D3 polypeptide 18kDa	0.55	0.37	-0.48	P	N	0.00	0.23	0.00	0.99	U	
Coja16207_c0_seq1	*00000002108		0.55	0.36	-0.01	P	N				1.01	FG	FG
Coja4948_c0_seq1	No hit	Novel	0.54	3.39	-0.22	P	N					U	
Coja4769_c0_seq1	*00000006885	GTPase HRas	0.54	0.96	-0.02	P	N	0.00	0.06	0.00	0.99	U	L
Coja9116_c0_seq1	*00000024468*	*CD59 molecule, complement regulatory protein*	0.54	0.58	0.22	P	Y	0.03	0.27	0.09	1.00	U	
Coja16921_c0_seq1	*00000012414	Glucosamine-phosphate N-acetyltransferase 1	0.54	0.42	**1.96**	P, GE	N	0.00	0.19	0.01	0.99	U	
Coja8306_c0_seq1	Not ortholog		0.54	0.11	0.55	P	N				0.52	U	
Coja4700_c0_seq1	*00000003046	Proteasome subunit alpha type	0.53	0.80	0.26	P	N	0.00	0.22	0.02	0.90	U	
Coja10754_c0_seq1	*00000010244	Aldo-keto reductase family 1, member A1 (aldehyde reductase)	0.53	0.72	**0.62**	P, GE	N	0.02	0.13	0.17	0.99	U	
Coja9023_c0_seq1	No hit	Novel	0.53	0.61	-0.13	P	N					U	
Coja10010_c0_seq1	*00000016536	Mitochondrial trifunctional protein, alpha subunit	0.53	0.38	**-0.81**	P	N	0.03	0.13	0.23	1.00	U	
Coja8189_c0_seq1	*00000000470	Lamin B2	0.53	0.21	0.42	P	N	0.01	0.14	0.04	0.98	U	FG

Coja20814_c0_seq1	**00000016623*	*Haloacid dehalogenase-like hydrolase domain containing 1*	0.53	0.09	0.22	P	N	0.07	0.15	0.49	1.00	U	T
Coja21605_c0_seq1	*00000004809	Peptidylprolyl isomerase F	0.53	0.08	0.18	P	N	0.01	0.09	0.14	1.00	U	L
Coja18012_c0_seq1	*00000010077	Uncharacterized protein	0.52	11.11	-0.14	P	N	0.00	0.16	0.00	1.00	U	
Coja25953_c0_seq1	*00000015379	Transgelin 3	0.52	2.37	-0.42	P	N				1.00	U	
Coja8951_c0_seq1	*00000024035	EPS8-like 2	0.52	0.58	0.30	P	N	0.01	0.16	0.03	0.99	U	FG
Coja5172_c0_seq1	*00000001162	Prohibitin	0.52	0.53	-0.55	P	N	0.00	0.14	0.00	0.99	U	L
Coja34842_c0_seq1	Not ortholog		0.52	0.31	**-1.25**	P	N				0.57	U	
Coja28654_c0_seq1	No hit	Novel	0.51	0.24	0.38	P	N					U	L
Coja12496_c0_seq1	**00000003797*	*Uncharacterized protein*	0.51	0.13	**-0.62**	P	N				1.00	U	L
Coja12366_c0_seq1	No hit	Novel	0.50	5.20	-0.32	P	N					U	
Coja12414_c0_seq1	*00000016144	Uncharacterized protein	0.50	1.90	**2.30**	P, GE	N	0.05	0.13	0.38	0.98	M	FG
Coja57514_c0_seq1	*00000000761	Tsukushin	0.50	0.81	**2.35**	P, GE	Y	0.04	0.19	0.20	0.90	M	FG
Coja3048_c0_seq1	*00000005805	Phospholipase C, delta 1	0.50	0.41	**1.18**	P, GE	N	0.01	0.08	0.13	1.00	U	FG
Coja98188_c0_seq1	Not ortholog		0.50	0.27	-0.78	P	N				0.29	U	
Coja9565_c0_seq1	*00000015764	Fatty acid binding protein 5	0.50	0.22	**-1.85**	P	N	0.01	0.06	0.17	0.99	U	
Coja3166_c0_seq1	*00000008455	Sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3C	0.50	0.20	0.93	P	Y	0.03	0.14	0.22	0.99	M	FG
Coja387812_c0_seq1	Not ortholog		0.50	0.00		P	N						
Coja4893_c0_seq1	*00000000489		0.49	2.81	-0.23	P	N				1.00	U	
Coja19052_c0_seq1	*00000002286	Histone H3.3	0.49	2.17	-0.44	P	N				0.99	U	
Coja10696_c0_seq1	*00000016241	ATPase, H+ transporting, lysosomal accessory protein 2	0.49	0.83	**0.85**	P, GE	Y	0.02	0.15	0.13	1.00	U	
Coja9955_c0_seq1	*00000016991	Esterase D	0.49	0.45	**-0.48**	P	N	0.06	0.29	0.21	1.00	U	L

Coja14999_c0_seq1	**00000010271*	*Glutathione reductase*	0.49	0.31	0.08	P	N	0.18	0.43	0.42	1.12	U	
Coja203399_c0_seq1	*00000006919	Premature ovarian failure, 1B	0.49	0.12	-0.06	P	N	0.04	0.27	0.13	0.30	U	FG
Coja17642_c0_seq1	*00000003957	Apolipoprotein H (beta-2-glycoprotein I)	0.49	0.11	-0.49	P	Y	0.03	0.14	0.23	0.98	U	L
Coja18142_c0_seq1	*00000002837	6S ribosomal protein L27	0.48	8.19	-0.52	P	N	0.00	0.37	0.00	0.99	U	
Coja41373_c0_seq1	Not ortholog		0.48	1.11	**2.41**	P, GE	N				0.23	U	FG
Coja16991_c0_seq1	*00000004719	Spectrin, alpha, non-erythrocytic 1 (alpha-fodrin)	0.48	0.43	0.33	P	N	0.00	0.13	0.02	1.00	U	
Coja12390_c0_seq1	*00000013708	Cytochrome b5	0.48	0.41	0.02	P	N	0.01	0.16	0.07	0.99	U	L
Coja17181_c1_seq1	*00000008156	Spectrin, beta, non-erythrocytic 1	0.48	0.40	-0.29	P	N	0.00	0.11	0.03	1.00	U	
Coja3454_c0_seq1	*00000002041	Agrin	0.48	0.20	**0.98**	P, GE	N	0.01	0.23	0.06	0.60	U	
Coja3430_c0_seq1	**00000004218*	*Uncharacterized protein*	0.48	0.09	0.35	P	N	0.01	0.08	0.14	0.99	U	U
Coja22428_c0_seq1	*00000002448	T-complex protein 1 subunit zeta	0.47	0.77	0.08	P	N	0.00	0.15	0.03	1.00	U	
Coja7414_c0_seq1	*00000002948	Thioredoxin-like 1	0.47	0.52	-0.16	P	N	0.00	0.17	0.00	1.00	U	
Coja17434_c0_seq1	*00000010175	Heat shock 9kDa protein 1, beta	0.47	0.27	**-0.64**	P	N	0.01	0.18	0.06	0.62	U	
Coja38973_c0_seq1	*00000014662	Retinol binding protein 5, cellular	0.47	0.14	**-1.06**	P	N	0.01	0.12	0.06	0.99	U	L
Coja17875_c0_seq1	No hit	Novel	0.47	0.01		P	N					T	T
Coja12319_c0_seq1	Not ortholog		0.46	0.53	-0.03	P	N				0.98	FG	FG
Coja5114_c0_seq1	*00000019759	Mitogen-activated protein kinase 14	0.46	0.45	-0.03	P	N	0.00	0.10	0.00	1.00	U	
Coja17257_c0_seq1	*00000005678	Filamin B, beta (actin binding protein 278)	0.46	0.44	0.61	P	N	0.00	0.11	0.02	1.00	U	FG
Coja9295_c0_seq1	*00000002064	Stomatin (EPB72)-like 2	0.46	0.36	-0.40	P	N	0.25	0.40	0.63	1.03	U	
Coja9161_c0_seq1	*00000005723		0.46	0.28	-0.05	P	N				1.00	U	
Coja14901_c0_seq1	Not ortholog		0.46	0.18	**-1.88**	P	N				0.96	U	L

Coja10270_c0_seq1	*00000004293	Uncharacterized protein	0.46	0.01	-0.31	P	Y	0.02	0.11	0.22	0.99	M	L
Coja4839_c0_seq1	*00000004588	Ribosomal protein S14	0.45	5.19	-0.49	P	N	0.00	0.17	0.00	0.92	U	
Coja10317_c0_seq1	*00000011511	Creatine kinase B-type	0.45	1.41	**1.04**	P, GE	N	0.02	0.10	0.19	0.98	U	T
Coja31077_c0_seq1	*00000003912	Adenylyl cyclase-associated protein	0.45	0.68	0.56	P	N	0.05	0.17	0.29	1.00	U	
Coja12874_c0_seq1	**00000007987*	*Hydroxysteroid (17-beta) dehydrogenase 12*	0.45	0.62	**0.83**	P, GE	N	0.21	0.37	0.55	1.00	U	L
Coja15200_c0_seq1	*00000008086	ARP1 actin-related protein 1 homolog A, centractin alpha	0.45	0.40	0.25	P	N	0.00	0.10	0.00	1.00	U	
Coja13819_c0_seq1	*00000017153	Radixin	0.45	0.37	-0.02	P	N	0.00	0.09	0.00	1.00	U	
Coja14390_c0_seq1	*00000016325	Glutathione S-transferase	0.45	0.34	0.03	P	N	0.05	0.14	0.38	1.00	U	L
Coja4493_c0_seq1	*00000016918	Uncharacterized protein	0.45	0.31	**-1.50**	P	N				1.00	U	
Coja11616_c0_seq1	Not ortholog		0.45	0.17	0.10	P	N				0.42	U	
Coja1061359_c0_seq1	No hit	Novel	0.45	0.00		P	N						
Coja14492_c0_seq1	*00000006565	Uncharacterized protein	0.44	1.67	**3.42**	P, GE	N				0.95	M	FG
Coja8622_c0_seq1	Not ortholog		0.44	0.69	-0.06	P	N				0.45	U	
Coja11815_c0_seq1	**00000000416*		0.44	0.64	-0.09	P	N				0.59	U	FG
Coja396518_c0_seq1	*00000017398	Uncharacterized protein	0.44	0.13	0.51	P	Y						
Coja18356_c0_seq1	*00000011523	Uncharacterized protein	0.43	6.12	-0.34	P	N	0.00	0.19	0.02	0.93	U	
Coja15098_c0_seq1	*00000006147	Nucleobindin 2	0.43	1.48	**2.50**	P, GE	Y	0.01	0.13	0.05	1.00	U	
Coja4745_c0_seq1	*00000011174	Proteasome (prosome, macropain) subunit, beta type, 1	0.43	0.76	-0.17	P	N	0.02	0.24	0.07	1.00	U	
Coja60626_c0_seq1	Not ortholog		0.43	0.66	**5.17**	P, GE	N				0.29	U	FG
Coja5155_c0_seq1	*00000005980		0.43	0.66	0.43	P	N				0.93	U	
Coja11900_c0_seq2	*00000015663	Hydroxysteroid dehydrogenase like 2	0.43	0.44	-0.29	P	N	0.02	0.10	0.21	1.01	U	L

Coja7824_c0_seq1	*00000005842	Glycerophosphodiester phosphodiesterase domain containing 2	0.43	0.34	-0.07	P	N	0.01	0.20	0.07	1.00	M	FG
Coja128928_c0_seq1	Not ortholog		0.43	0.13	0.17	P	N				0.20	U	
Coja776532_c0_seq1	No hit	Novel	0.43	0.00		P	N						
Coja5091_c0_seq1	*00000012726	Heat shock protein 9kDa beta (Grp94), member 1	0.42	5.02	**1.52**	P, GE	Y	0.00	0.12	0.02	0.96	U	U
Coja4471_c0_seq1	*00000002447	Catenin (cadherin-associated protein), alpha 1, 12kDa	0.42	0.79	0.23	P	N	0.00	0.12	0.00	1.00	U	
Coja14552_c0_seq1	Not ortholog		0.42	0.33	-0.41	P	N				0.68	U	
Coja10441_c0_seq1	*00000015331	UDP-N-acetylglucosamine-2-epimerase/N-acetylmannosamine kinase	0.41	4.56	**2.18**	P, GE	N	0.01	0.12	0.05	1.00	U	FG
Coja5139_c0_seq1	*00000016140	Family with sequence similarity 3, member B	0.41	2.89	**2.01**	P, GE	N	0.05	0.09	0.51	0.97	M	FG
Coja5959_c0_seq1	*00000011404	Homeodomain-only protein	0.41	1.05	-0.28	P	N	0.01	0.16	0.04	0.99	U	FG
Coja5030_c0_seq1	*00000008600	IQ motif containing GTPase activating protein 1	0.41	0.92	**0.94**	P, GE	N	0.01	0.11	0.10	1.00	U	FG
Coja3605_c0_seq1	Not ortholog		0.41	0.72	0.17	P	Y				0.92	FG	FG
Coja5728_c0_seq1	*00000003015	Uncharacterized protein	0.41	0.25	**-1.56**	P	Y	0.01	0.06	0.21	1.00	U	L
Coja9170_c0_seq1	*00000008409	Uncharacterized protein	0.41	0.21	-0.31	P	N	0.03	0.25	0.14	1.00	U	L
Coja18465_c0_seq1	*00000017299	Ribosomal protein S11	0.40	7.00	-0.40	P	N	0.01	0.29	0.05	1.01	U	
Coja4880_c0_seq1	*00000016232	Uncharacterized protein	0.40	6.33	-0.08	P	N	0.00	0.24	0.00	0.95	U	

Coja24675_c0_seq1	*00000005129	*Proteasome (prosome, macropain) subunit, alpha type, 7*	0.40	0.48	**-0.72**	P	N	0.00	0.22	0.00	0.88	U	
Coja55982_c0_seq1	*00000015438	Myelin protein zero-like 1	0.40	0.38	0.54	P	Y	0.09	0.28	0.33	0.49	U	
Coja177736_c0_seq1	No hit	Novel	0.40	0.13	1.31	P	N					M	FG
Coja4900_c0_seq1	*00000015637	6S ribosomal protein L7	0.39	8.38	-0.28	P	N	0.00	0.09	0.02	0.97	U	
Coja11948_c0_seq1	*00000016666	Cathepsin B	0.39	1.94	0.23	P	Y	0.01	0.18	0.03	1.00	U	L
Coja26643_c0_seq1	*00000012039	Proteasome (prosome, macropain) subunit, alpha type, 3	0.39	0.60	-0.31	P	N	0.00	0.17	0.00	0.94	U	
Coja15763_c0_seq1	*00000014248	Transmembrane protein 33	0.39	0.31	**1.12**	P, GE	N	0.00	0.10	0.04	1.00	U	
Coja15385_c0_seq1	*00000011272	RAB5A, member RAS oncogene family	0.39	0.30	-0.08	P	N				1.00	U	
Coja12913_c0_seq1	*00000017005	Calcium binding protein 39-like	0.39	0.30	-0.22	P	N	0.00	0.11	0.00	0.99	M	FG
Coja5982_c0_seq1	*00000007261*	*Epithelial membrane protein 2*	0.38	1.18	**0.77**	P, GE	Y	0.02	0.05	0.31	0.92	U	FG
Coja14674_c0_seq1	*00000013702	PREDICTED: aminopeptidase puromycin sensitive	0.38	0.57	0.31	P	N				1.00	U	
Coja15382_c0_seq1	*00000019717*	*Uncharacterized protein*	0.38	0.33	0.27	P	N	0.04	0.25	0.15	1.01	M	FG
Coja15195_c0_seq1	Not ortholog		0.38	0.07	-1.30	P	N				0.84	M	L
Coja17940_c0_seq1	*00000016977*	*Translationally-controlled tumor protein homolog*	0.37	10.55	-0.80	P	N	0.00	0.12	0.03	0.99	U	
Coja4537_c0_seq1	*00000012843	GDP-mannose 4,6-dehydratase	0.37	1.15	**1.10**	P, GE	N	0.00	0.07	0.00	0.99	U	FG
Coja12508_c0_seq1	*00000005139	Clathrin, heavy chain (Hc)	0.37	0.82	**1.40**	P, GE	N	0.00	0.09	0.00	0.99	U	
Coja10476_c0_seq1	Not ortholog		0.37	0.67	**-3.45**	P	Y				0.88	U	L
Coja27449_c0_seq1	*00000021410*	*Uncharacterized protein*	0.37	0.43	**1.21**	P, GE	N	0.12	0.40	0.30	0.92	U	

Coja238139_c0_seq1	No hit	Novel	0.37	0.00		P	N					T	T
Coja18573_c0_seq1	**00000003954*	*Ribosomal protein L11*	0.36	5.15	-0.08	P	N	0.00	0.47	0.00	0.99	U	
Coja15060_c0_seq1	*00000011651	Uncharacterized protein	0.36	2.72	**1.13**	P, GE	N	0.00	0.17	0.00	0.99	U	
Coja5628_c0_seq1	*00000009546	HtrA serine peptidase 1	0.36	1.68	**1.79**	P, GE	Y	0.01	0.10	0.07	1.00	U	FG
Coja12886_c0_seq1	*00000008844	Solute carrier family 31 (copper transporters), member 1	0.36	1.02	**1.36**	P, GE	N	0.01	0.14	0.07	0.94	U	
Coja13797_c0_seq1	*00000006904	Ribonuclease/angiogenin inhibitor 1	0.36	0.39	0.29	P	N	0.01	0.13	0.05	1.00	U	
Coja14798_c0_seq1	**00000024317*	*Chromosome 15 open reading frame 48*	0.36	0.25	-0.38	P	N	0.04	0.19	0.20	1.00	M	FG
Coja19027_c0_seq1	*00000009829	Ribosomal protein L18a	0.35	4.88	-0.11	P	N	0.03	0.43	0.06	0.95	U	
Coja26527_c0_seq1	*00000007253	Interleukin 1 receptor accessory protein	0.35	1.14	**2.72**	P, GE	Y	0.03	0.18	0.19	0.59	U	M
Coja4398_c0_seq1	Not ortholog		0.35	1.07	**0.97**	P, GE	N				0.19	U	FG
Coja20733_c0_seq1	*00000005789	Uncharacterized protein	0.35	0.91	-0.20	P	N	0.02	0.14	0.13	1.00	U	
Coja23791_c0_seq1	*00000011950	Aconitase 2, mitochondrial	0.35	0.67	-0.68	P	N	0.00	0.10	0.02	1.00	U	
Coja12072_c0_seq1	*00000006015	Coatomer subunit beta	0.35	0.63	**1.28**	P, GE	N	0.00	0.14	0.00	1.00	U	
Coja15031_c0_seq1	*00000005654	Phosphatidylinositol transfer protein, beta	0.35	0.57	**0.66**	P, GE	N	0.01	0.20	0.06	0.99	U	
Coja5165_c0_seq1	*00000006198	Lanosterol synthase (2,3-oxidosqualene-lanosterol cyclase)	0.35	0.22	0.44	P	N	0.04	0.23	0.17	0.98	U	L
Coja17109_c0_seq1	*00000013848	Uncharacterized protein	0.35	0.13	0.25	P	N				1.01	U	
Coja557001_c0_seq1	No hit	Novel	0.35	0.04	0.14	P	N						

Coja16773_c0_seq1	*00000000085	Glucosamine-fructose-6-phosphate aminotransferase	0.34	1.50	**1.91**	P, GE	N	0.00	0.12	0.01	1.00	U	FG
Coja23063_c0_seq1	Not ortholog		0.34	1.43	**0.62**	P, GE	Y				0.93	U	L
Coja12836_c0_seq1	Not ortholog		0.34	1.07	-0.51	P	N				0.69	U	FG
Coja20750_c0_seq1	Not ortholog		0.34	1.04	-0.28	P	N				0.96	U	
Coja13981_c0_seq1	*00000015436	Syntenin	0.34	0.98	**0.73**	P, GE	N	0.01	0.19	0.07	1.00	U	
Coja21277_c0_seq1	*00000016336	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 9, 22kDa	0.34	0.65	-0.54	P	N	0.01	0.28	0.04	0.96	U	L
Coja4533_c0_seq1	Not ortholog		0.34	0.35	0.57	P	N				0.38	U	
Coja11957_c0_seq1	*00000004763	Aldehyde dehydrogenase 3 family, member A2	0.34	0.33	**0.55**	P, GE	N	0.02	0.25	0.10	1.00	U	L
Coja160991_c0_seq1	*00000002905*	*Myeloid-associated differentiation marker-like 2*	0.34	0.15	-0.46	P	N				0.97	U	
Coja54906_c0_seq1	*00000000161	Isochorismatase domain containing 1	0.34	0.12	-0.11	P	N	0.00	0.09	0.02	0.93	U	
Coja17868_c0_seq1	*00000015617	Uncharacterized protein	0.33	8.13	-0.35	P	N	0.02	0.28	0.06	0.97	U	
Coja18292_c0_seq1	*00000011951	4S ribosomal protein SA	0.33	7.12	0.08	P	N	0.00	0.11	0.00	1.00	U	
Coja19892_c0_seq1	*00000006753	ATP synthase, H+ transporting, mitochondrial F1 complex, gamma polypeptide 1	0.33	1.16	-0.29	P	N	0.01	0.11	0.06	1.00	U	
Coja22446_c0_seq1	Not ortholog		0.33	1.04	0.00	P	N				0.71	U	U
Coja26807_c0_seq1	*00000005163	Mitochondrial carrier homolog 2	0.33	0.81	0.20	P	N	0.00	0.20	0.02	0.99	U	
Coja12372_c0_seq1	*00000011331	Uncharacterized protein	0.33	0.76	**1.31**	P, GE	N	0.05	0.16	0.28	1.00	U	L
Coja19993_c0_seq1	*00000013992	Uncharacterized protein	0.33	0.42	0.66	P	Y				1.00	U	L

Coja111035_c0_seq1	*00000006995	Apolipoprotein D	0.33	0.24	2.35	P	Y	0.06	0.28	0.22	0.99	FG	FG
Coja49839_c0_seq1	Not ortholog		0.33	0.22	**0.82**	P, GE	N				0.24	U	
Coja263195_c0_seq1	Not ortholog		0.33	0.11	0.67	P	N						
Coja4903_c0_seq1	*00000004818	Ribosomal protein L6	0.32	7.83	-0.23	P	N	0.03	0.18	0.16	0.87	U	
Coja12400_c0_seq1	*00000005945	Calnexin	0.32	1.77	**1.22**	P, GE	Y	0.02	0.15	0.15	0.87	U	
Coja21602_c0_seq1	*00000007971	Small ubiquitin-related modifier 2	0.32	1.23	-0.09	P	N	0.00	0.08	0.00	0.99	U	
Coja18802_c0_seq1	*00000008130	Reticulon 4	0.32	0.93	**-1.24**	P	N	0.01	0.08	0.08	1.00	U	
Coja15868_c0_seq1	*00000015416	Similar to Zgc:63829	0.32	0.48	**0.78**	P, GE	N	0.03	0.12	0.27	1.00	U	M
Coja17173_c0_seq1	Not ortholog		0.32	0.35	**1.54**	P, GE	N				0.55	M	FG
Coja9006_c0_seq1	*00000003250	Uncharacterized protein	0.32	0.35	-0.02	P	N	0.01	0.20	0.05	0.96	U	
Coja13399_c0_seq1	*00000007315	Proteasome (prosome, macropain) 26S subunit, non-ATPase, 6	0.32	0.32	-0.15	P	N	0.00	0.34	0.01	0.97	U	
Coja17412_c0_seq1	*00000014122	Uncharacterized protein	0.32	0.27	**-0.83**	P	N	0.01	0.13	0.10	1.00	U	
Coja13274_c0_seq1	*00000008475	Arrestin domain containing 1	0.32	0.24	**0.74**	P, GE	N	0.01	0.12	0.06	0.97	U	
Coja11965_c0_seq1	Not ortholog		0.32	0.17	0.65	P	N				0.44	U	L
Coja5422_c0_seq1	*00000014682	Calpastatin	0.32	0.16	-0.44	P	N	0.09	0.22	0.40	0.59	U	T
Coja10532_c0_seq1	No hit	Novel	0.32	0.05	**1.38**	P, GE	N					M	
Coja22030_c0_seq1	*00000005402	Dynein, light chain, roadblock-type 2	0.32	0.01		P	N	0.02	0.10	0.23	0.99	T	T
Coja19602_c0_seq1	*00000000719	Ribosomal protein L22	0.31	1.85	**-0.84**	P	N	0.01	0.21	0.05	0.81	U	
Coja10289_c0_seq1	Not ortholog		0.31	0.90	-0.35	P	N				0.39	U	L
Coja26546_c0_seq1	*00000002373*	*Uncharacterized protein*	0.31	0.68	-0.18	P	N	0.01	0.22	0.04	0.95	U	
Coja5806_c0_seq1	*00000024112	Proteasome (prosome, macropain) 26S subunit, non-ATPase, 11	0.31	0.44	-0.31	P	N				1.00	U	

Coja3201_c0_seq1	*00000000299	Plakophilin 1 (ectodermal dysplasia/skin fragility syndrome)	0.31	0.32	0.41	P	N	0.00	0.15	0.02	1.00	M	FG
Coja15934_c0_seq1	*00000012186	Caspase 6	0.31	0.27	0.61	P	N	0.02	0.14	0.17	1.00	U	FG
Coja17646_c0_seq1	*00000006629	Retinol-binding protein 4	0.31	0.17	-0.34	P	Y	0.00	0.12	0.04	0.99	U	L
Coja350043_c0_seq1	Not ortholog		0.31	0.00		P	N						
Coja4843_c0_seq1	Not ortholog		0.30	7.27	0.75	P	Y				0.84	U	
Coja4868_c0_seq1	*00000001658	Ribosomal protein L19	0.30	5.59	-0.09	P	N	0.00	0.23	0.00	0.99	U	
Coja18862_c0_seq1	*00000011290	Ribosomal protein L15	0.30	4.97	-0.23	P	N	0.00	0.11	0.00	0.99	U	
Coja18372_c0_seq1	*00000016172	6S acidic ribosomal protein P1	0.30	4.49	-0.38	P	N	0.00	0.16	0.00	0.99	U	
Coja14233_c0_seq1	*00000015256	Peptidylglycine alpha-amidating monooxygenase	0.30	1.33	**1.87**	P, GE	Y	0.03	0.20	0.17	0.48	U	FG
Coja16749_c0_seq1	*00000012298	Solute carrier family 39 (zinc transporter), member 8	0.30	0.93	**2.06**	P, GE	N	0.01	0.09	0.11	0.97	M	FG
Coja13621_c0_seq1	*00000005776		0.30	0.49	0.12	P	N				0.99	U	
Coja21533_c0_seq1	*00000009086	EH-domain containing 3	0.30	0.48	0.24	P	N	0.00	0.08	0.02	1.00	U	T
Coja16543_c0_seq1	*00000005077	Coagulation factor VIII, procoagulant component	0.30	0.37	**1.37**	P, GE	N	0.05	0.13	0.35	0.85	U	FG
Coja14512_c0_seq1	*00000005998	Related RAS viral (r-ras) oncogene homolog 2	0.30	0.36	**0.52**	P, GE	N	0.00	0.12	0.00	1.00	U	
Coja16571_c0_seq1	*00000016736	Cytoplasmic FMR1 interacting protein 1	0.30	0.35	0.48	P	N	0.00	0.11	0.00	1.00	U	
Coja16153_c0_seq1	Not ortholog		0.30	0.20	-0.04	P	N				0.64	U	
Coja704359_c0_seq1	Not ortholog		0.30	0.04	-1.64	P	N						

Coja22747_c0_seq1	*00000000707	Proteasome (prosome, macropain) 26S subunit, non-ATPase, 7	0.29	0.79	-0.39	P	N	0.00	0.10	0.00	0.87	U	
Coja10788_c0_seq1	*00000006294	Proteasome (prosome, macropain) 26S subunit, non-ATPase, 2	0.29	0.43	-0.02	P	N	0.01	0.25	0.02	0.98	U	
Coja9151_c0_seq1	*00000008869*	*Delta-aminolevulinic acid dehydratase*	0.29	0.29	**0.89**	P, GE	N	0.02	0.21	0.11	0.98	U	
Coja15191_c0_seq1	*00000001688	Septin 5	0.29	0.20	0.18	P	N	0.00	0.15	0.00	0.78	M	FG
Coja9459_c0_seq1	No hit	Novel	0.29	0.10	**-1.45**	P	N					U	
Coja286928_c0_seq1	Not ortholog		0.29	0.04	-0.40	P	N				0.86	U	U
Coja18493_c0_seq1	*00000007205	Cytochrome c oxidase subunit 6A, mitochondrial	0.28	1.91	-0.04	P	N	0.02	0.24	0.07	0.99	U	L
Coja22975_c0_seq1	Not ortholog		0.28	1.32	0.04	P	N				0.22	U	U
Coja5781_c0_seq1	*00000009651	Latexin	0.28	0.95	**2.10**	P, GE	N	0.06	0.15	0.41	0.98	U	FG
Coja15124_c0_seq1	*00000014708	Glutathione S-transferase kappa 1	0.28	0.94	**0.92**	P, GE	N	0.06	0.25	0.25	0.99	U	L
Coja21545_c0_seq1	*00000015849	Malic enzyme 1, NADP(+)-dependent, cytosolic	0.28	0.28	**0.82**	P, GE	N	0.01	0.13	0.10	1.00	U	L
Coja232154_c0_seq1	*00000010567	Uncharacterized protein	0.28	0.10	**1.61**	P, GE	N	0.01	0.15	0.05	1.00	M	FG
Coja13738_c0_seq1	*00000002581	HIV-1 Tat interactive protein 2, 3kDa	0.28	0.09	-0.59	P	N	0.05	0.28	0.19	0.95	U	
Coja468100_c0_seq1	Not ortholog		0.28	0.06	0.87	P	Y						
Coja635626_c0_seq1	No hit	Novel	0.28	0.05	0.30	P	Y						
Coja10482_c0_seq1	*00000005907	Ribophorin I	0.27	3.94	**1.48**	P, GE	Y	0.01	0.17	0.08	0.95	U	
Coja14943_c0_seq1	*00000001885	Uroplakin 3B-like	0.27	1.32	-0.30	P	Y	0.13	0.43	0.29	0.95	M	FG
Coja8991_c0_seq1	*00000005422		0.27	0.76	**-0.71**	P	N				1.00	U	FG
Coja15780_c0_seq1	Not ortholog		0.27	0.67	0.61	P	N				0.97	U	

Coja33251_c0_seq1	*00000013036	Vacuolar H+ ATPase E1	0.27	0.46	-0.48	P	N	0.01	0.07	0.08	1.00	U	
Coja6008_c0_seq1	*00000002187	Hydroxysteroid (17-beta) dehydrogenase 4	0.27	0.43	**1.04**	P, GE	N	0.07	0.20	0.33	0.80	U	L
Coja14704_c0_seq1	*00000010186	Calpain-1 catalytic subunit	0.27	0.32	-0.21	P	N	0.01	0.14	0.04	1.00	U	
Coja14354_c0_seq1	*00000007196	Syntaxin binding protein 1	0.27	0.24	0.28	P	N	0.01	0.30	0.05	1.00	U	
Coja12618_c0_seq1	*00000010303	Uncharacterized protein	0.26	0.86	0.65	P	N				1.00	U	L
Coja47782_c0_seq1	*00000007549	VAMP (vesicle-associated membrane protein)-associated protein B and C	0.26	0.69	0.09	P	N	0.00	0.05	0.10	0.91	U	
Coja5304_c0_seq1	*00000001428	EF hand domain family, member D1	0.26	0.68	**1.63**	P, GE	N	0.01	0.16	0.07	1.00	U	FG
Coja14574_c0_seq1	*00000002623	Procollagen C-endopeptidase enhancer 2	0.26	0.44	0.72	P	Y				0.86	U	
Coja11096_c0_seq1	No hit	Novel	0.26	0.42	0.08	P	N					U	
Coja5019_c0_seq1	*00000009899	Cullin-associated and neddylation-dissociated 1	0.26	0.30	**0.84**	P, GE	N	0.00	0.05	0.00	1.00	U	
Coja16447_c0_seq1	*00000014800	ATPase, H+ transporting, lysosomal 7kDa, V1 subunit A	0.26	0.15	**1.03**	P, GE	N	0.00	0.08	0.02	1.00	U	
Coja289758_c0_seq1	Not ortholog		0.26	0.13	-0.59	P	N				0.88	M	FG
Coja18052_c0_seq1	**00000005921*	*12K serum protein, beta-2-m cross-reactiveUncharacterized protein*	0.26	0.05	-0.35	P	Y				1.00	U	L
Coja18510_c0_seq1	*00000007699	Ribosomal protein S25	0.25	6.08	-0.27	P	N	0.01	0.38	0.03	0.74	U	

Coja12932_c0_seq1	*00000012415	Protein disulfide isomerase family A, member 4	0.25	1.67	**1.35**	P, GE	Y	0.02	0.08	0.27	0.98	U	
Coja16574_c0_seq1	*00000001501	Mitogen-activated protein kinase 1	0.25	0.74	**1.21**	P, GE	N	0.00	0.10	0.02	0.99	U	
Coja4718_c0_seq1	*00000002777	Acetyl-Coenzyme A acyltransferase 2 (mitochondrial 3-oxoacyl-Coenzyme A thiolase)	0.25	0.57	-0.27	P	N	0.01	0.11	0.11	1.00	U	L
Coja13601_c0_seq1	*00000007326	Phosphomannomutase 2	0.25	0.42	**1.75**	P, GE	N	0.01	0.22	0.04	0.98	U	FG
Coja9481_c0_seq1	*00000010187	Solute carrier family 1 (neuronal/epithelial high affinity glutamate transporter, system Xag), member 1	0.25	0.36	**1.10**	P, GE	N	0.02	0.12	0.14	1.00	M	FG
Coja6111_c0_seq1	*00000008009	B-cell receptor-associated protein 29	0.25	0.35	0.06	P	N	0.03	0.16	0.18	1.00	U	
Coja14715_c0_seq2	*00000023936	Uncharacterized protein	0.25	0.30	1.25	P	N	0.01	0.15	0.04	1.00	U	
Coja10339_c0_seq1	*00000021450*	*Uncharacterized protein*	0.25	0.04	-0.03	P	N				0.98	U	L
Coja14865_c0_seq1	*00000016088	T-complex protein 1 subunit eta	0.24	0.67	-0.17	P	N	0.14	0.38	0.38	0.98	U	
Coja10443_c0_seq1	*00000013041	Chaperonin containing TCP1, subunit 5 (epsilon)	0.24	0.65	-0.15	P	N	0.00	0.17	0.01	0.99	U	
Coja4558_c0_seq1	*00000012784	Thioredoxin domain containing 5	0.24	0.65	**0.72**	P, GE	Y	0.02	0.11	0.16	0.61	U	FG
Coja4936_c0_seq1	*00000017163	Uncharacterized protein	0.24	0.38	**-0.65**	P	N	0.02	0.19	0.11	0.98	U	L

Coja12994_c0_seq1	*00000005398	Hydroxyacylglutathione hydrolase	0.24	0.25	-0.11	P	N	0.01	0.12	0.06	0.91	U	
Coja54009_c0_seq1	*00000007509	PREDICTED: similar to MGC68706 protein isoform 1	0.24	0.23	0.08	P	N				1.00	U	
Coja105197_c0_seq1	Not ortholog		0.24	0.18	0.27	P	N				0.54	U	
Coja15649_c0_seq1	*00000015881	Copine III	0.24	0.18	0.60	P	N	0.01	0.10	0.09	0.99	M	FG
Coja9079_c0_seq1	Not ortholog		0.24	0.15	0.17	P	N				0.54	U	
Coja236907_c0_seq1	Not ortholog		0.24	0.03	0.41	P	N				0.12	T	T
Coja19990_c0_seq1	Not ortholog		0.24	0.02	**-0.89**	P	N				0.97	U	T
Coja5457_c0_seq1	*00000011560	PARK2 co-regulated	0.24	0.00		P	N	0.04	0.09	0.40	1.00	T	T
Coja463296_c0_seq1	Not ortholog		0.24	0.00		P	N						
Coja402108_c0_seq1	Not ortholog		0.24	0.00		P	N				0.22	T	T
Coja4947_c0_seq1	*00000008671	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 6	0.23	2.43	**2.98**	P, GE	N	0.11	0.22	0.50	0.93	U	FG
Coja18567_c0_seq1	Not ortholog		0.23	2.20	-0.08	P	N				0.42	U	L
Coja8434_c0_seq1	*00000012420	Beta-galactoside-binding lectin	0.23	0.57	**-1.85**	P	N	0.02	0.13	0.19	0.99	U	FG
Coja17112_c0_seq1	*00000007215	Uncharacterized protein	0.23	0.42	**0.68**	P, GE	N	0.00	0.14	0.00	1.00	U	
Coja16082_c0_seq1	*00000015777*	*Inositol(myo)-1(or 4)-monophosphatase 1*	0.23	0.20	**0.87**	P, GE	N	0.00	0.12	0.04	1.00	U	
Coja17261_c0_seq1	*00000011805	Uncharacterized protein	0.23	0.19	0.37	P	Y	0.04	0.13	0.27	1.00	U	L
Coja8128_c0_seq1	No hit	Novel	0.23	0.11	-0.38	P	N					U	
Coja353138_c0_seq1	*00000008831*	*Galectin-related protein*	0.23	0.08	-0.40	P	N	0.00	0.12	0.03	0.76	U	FG
Coja21055_c0_seq1	Not ortholog		0.23	0.03	-0.51	P	N				0.58	U	L
Coja18600_c0_seq1	*00000002157	Uncharacterized protein	0.22	5.41	-0.58	P	N				0.99	U	

Coja12308_c0_seq1	*00000000054	Ornithine aminotransferase	0.22	2.82	**1.41**	P, GE	N	0.01	0.16	0.05	1.00	U	
Coja15874_c0_seq1	*00000000608	Hypothetical protein LOC41586	0.22	1.56	0.86	P	Y	0.07	0.35	0.21	1.00	U	M
Coja10874_c0_seq1	*00000005869	Coatomer protein complex, subunit gamma	0.22	1.23	**1.43**	P, GE	N	0.00	0.14	0.00	1.00	U	
Coja22588_c0_seq1	No hit	Novel	0.22	0.90	-0.25	P	N					U	
Coja17571_c0_seq1	*00000011330	Uncharacterized protein	0.22	0.40	0.12	P	N	0.00	0.13	0.04	1.00	U	
Coja10051_c0_seq1	Not ortholog		0.22	0.31	-0.46	P	N				0.97	U	L
Coja17106_c0_seq1	*00000023424	Uncharacterized protein	0.22	0.29	**5.31**	P, GE	N				0.97	M	FG
Coja37229_c0_seq1	*00000010410	Profilin 2	0.22	0.20	-0.59	P	N	0.07	0.26	0.27	0.99	U	
Coja12858_c0_seq1	*00000010185	Heat shock 7kDa protein 4-like	0.22	0.20	0.58	P	N	0.01	0.10	0.09	1.00	U	T
Coja20490_c0_seq1	*00000015504	Tocopherol (alpha) transfer protein	0.22	0.14	0.30	P	N	0.03	0.10	0.31	0.98	U	L
Coja1698_c0_seq1	*00000012872	Uncharacterized protein	0.22	0.11	1.11	P	N	0.07	0.13	0.53	1.02	FG	FG
Coja206916_c0_seq1	Not ortholog		0.22	0.10	**5.35**	P, GE	Y				0.33	U	
Coja15550_c0_seq1	*00000005114	Uncharacterized protein	0.22	0.06	**-1.64**	P	Y				0.97	U	L
Coja4256_c0_seq1	*00000007382	Uncharacterized protein	0.22	0.02	0.24	P	N				0.71	M	L
Coja12405_c0_seq1	*00000023070	Uncharacterized protein	0.22	0.02	-0.32	P	Y	0.08	0.18	0.46	1.00	L	L
Coja12217_c0_seq1	No hit	Novel	0.21	9.04	-0.61	P	N					U	
Coja4809_c0_seq1	No hit	Novel	0.21	2.87	-0.47	P	N					U	
Coja10393_c0_seq1	*00000011905	Catenin (cadherin-associated protein), beta 1, 88kDa	0.21	1.52	0.51	P	N	0.00	0.08	0.00	1.00	U	
Coja5083_c0_seq1	Not ortholog		0.21	1.09	**0.72**	P, GE	N				0.43	U	
Coja5854_c0_seq1	*00000009039	Neural proliferation, differentiation and control, 1	0.21	0.62	**1.69**	P, GE	Y	0.03	0.25	0.11	1.00	U	FG
Coja22143_c0_seq1	*00000015821	Chaperonin containing TCP1, subunit 8 (theta)	0.21	0.53	-0.17	P	N	0.00	0.12	0.02	1.00	U	

Coja24976_c0_seq1	*00000004442	Pyrophosphatase (inorganic) 1	0.21	0.51	**0.56**	P, GE	N	0.01	0.23	0.04	0.94	U	L
Coja46610_c0_seq1	*00000016993	Succinate-CoA ligase, ADP-forming, beta subunit	0.21	0.36	0.06	P	N	0.01	0.07	0.08	1.00	U	
Coja16305_c0_seq1	*00000001242		0.21	0.33	0.13	P	Y				0.99	U	
Coja17158_c0_seq1	*00000006757	Xylosyltransferase I	0.21	0.26	**1.80**	P, GE	N	0.01	0.16	0.05	0.88	FG	FG
Coja15505_c0_seq1	*00000006723	Uncharacterized protein	0.21	0.26	0.62	P	N				0.98	U	L
Coja15903_c0_seq1	*00000002536	Syntaxin-2	0.21	0.24	**0.87**	P, GE	N	0.06	0.24	0.27	0.90	U	
Coja6153_c0_seq1	*00000007102	Malectin	0.21	0.23	**1.72**	P, GE	N	0.03	0.17	0.20	0.87	U	
Coja10703_c0_seq1	*00000006250	Trafficking protein particle complex 2-like	0.21	0.22	0.05	P	N	0.00	0.34	0.00	0.99	U	
Coja31154_c0_seq1	*00000009803	Microsomal glutathione S-transferase 2	0.21	0.21	0.00	P	Y	0.07	0.14	0.51	0.99	U	L
Coja4359_c0_seq1	Not ortholog		0.21	0.19	**-0.68**	P	N				0.95	U	
Coja4085_c0_seq1	No hit	Novel	0.21	0.14	0.50	P	N						
Coja18314_c0_seq1	*00000002868	6S ribosomal protein L26	0.20	6.58	-0.59	P	N				0.87	U	
Coja18434_c0_seq1	*00000015195	4S ribosomal protein S15	0.20	5.62	-0.36	P	N				0.94	U	
Coja4413_c0_seq1	*00000008424	Calmodulin-like 3	0.20	1.63	**0.90**	P, GE	N	0.00	0.38	0.00	0.96	M	FG
Coja21372_c0_seq1	*00000012311	Ubiquitin-conjugating enzyme E2D 3	0.20	1.36	0.31	P	N	0.00	0.03	0.00	0.99	U	U
Coja4924_c0_seq1	*00000009153	Coatomer protein complex, subunit alpha	0.20	0.96	**1.22**	P, GE	N	0.04	0.21	0.20	0.67	U	
Coja11088_c0_seq1	*00000007462	Uncharacterized protein	0.20	0.67	**0.66**	P, GE	Y	0.03	0.17	0.15	0.99	U	
Coja34235_c0_seq1	*00000010689	Protein mago nashi homolog	0.20	0.44	-0.38	P	N	0.00	0.37	0.00	0.99	U	
Coja4357_c0_seq1	*00000008015	TMEM189-UBE2V1 readthrough	0.20	0.37	0.01	P	N	0.00	0.06	0.00	0.99	U	
Coja37015_c0_seq1	Not ortholog		0.20	0.35	-0.09	P	N				0.51	U	

Coja13470_c0_seq1	*00000011128	Proteasome (prosome, macropain) 26S subunit, non-ATPase, 14	0.20	0.34	0.39	P	N	0.00	0.08	0.00	0.97	U	
Coja30357_c0_seq1	*00000007737	Uncharacterized protein	0.20	0.23	-0.22	P	N	0.05	0.22	0.21	0.98	U	L
Coja3154_c0_seq1	Not ortholog		0.20	0.19	**2.90**	P, GE	N				0.65	U	
Coja6975_c0_seq1	*00000017219	Chromosome 11 open reading frame 54	0.20	0.16	0.42	P	N	0.02	0.16	0.09	0.97	U	L
Coja16213_c0_seq1	*00000008534	Polymerase (RNA) II (DNA directed) polypeptide H	0.20	0.15	-0.48	P	N				0.99	U	
Coja490080_c0_seq1	*00000012893*	*Pleiotrophin*	0.20	0.08	**2.18**	P, GE	Y	0.00	0.12	0.00	0.78	M	FG
Coja4933_c0_seq1	*00000015339	Ribosomal protein L24	0.19	5.06	-0.30	P	N	0.00	0.11	0.00	0.99	U	
Coja5198_c0_seq1	*00000013155	Uncharacterized protein	0.19	2.44	0.39	P	N				1.01	U	
Coja24719_c0_seq1	*00000005357	Coatomer protein complex, subunit beta 2 (beta prime)	0.19	2.05	**1.90**	P, GE	N	0.00	0.16	0.03	0.67	U	
Coja31134_c0_seq1	*00000005551	Dickkopf homolog 3	0.19	1.69	**1.07**	P, GE	N	0.04	0.14	0.26	0.88	U	FG
Coja13391_c0_seq1	*00000016129	Tissue specific transplantation antigen P35B	0.19	0.98	**1.68**	P, GE	N	0.01	0.24	0.06	1.00	U	FG
Coja24374_c0_seq1	*00000001415	Ubiquitin-conjugating enzyme E2L 3	0.19	0.77	0.24	P	N	0.00	0.05	0.00	0.99	U	
Coja15399_c0_seq1	*00000005509	Dihydropyrimidine dehydrogenase	0.19	0.71	**1.76**	P, GE	N	0.03	0.14	0.21	1.00	U	M
Coja5843_c0_seq1	Not ortholog		0.19	0.57	-0.41	P	Y				0.99	U	
Coja6100_c0_seq1	Not ortholog		0.19	0.54	**1.29**	P, GE	N				0.06	U	FG
Coja4870_c0_seq1	*00000004245	Nudix (nucleoside diphosphate linked moiety X)-type motif 1	0.19	0.52	**1.42**	P, GE	N	0.06	0.20	0.29	0.91	U	L

Coja12773_c0_seq1	*00000004005	PREDICTED: similar to aflatoxin aldehyde reductase	0.19	0.34	0.25	P	N				0.87	U	L
Coja16841_c0_seq1	*00000004375	Vacuolar protein sorting 35	0.19	0.27	0.22	P	N	0.00	0.09	0.01	1.00	U	U
Coja135303_c0_seq1	Not ortholog		0.19	0.22	0.13	P	N				0.20	U	FG
Coja182787_c0_seq1	*00000004355	Mitochondrial calcium uniporter	0.19	0.19	-0.14	P	N	0.01	0.09	0.09	0.93	U	FG
Coja146439_c0_seq1	Not ortholog		0.19	0.17	0.87	P	N				0.18	U	
Coja9158_c0_seq1	*00000016313	Glutamate-cysteine ligase, catalytic subunit	0.19	0.16	-0.17	P	N	0.00	0.16	0.02	1.00	U	
Coja16624_c0_seq1	*00000008736	Patatin-like phospholipase domain containing 7	0.19	0.15	0.09	P	N	0.05	0.19	0.28	1.00	U	
Coja17344_c0_seq1	*00000002747	Fatty acid synthase	0.19	0.14	0.61	P	N	0.01	0.11	0.12	1.00	U	L
Coja45100_c0_seq1	Not ortholog		0.19	0.11	-0.26	P	N				0.96	U	L
Coja11826_c0_seq1	Not ortholog		0.19	0.02	**-2.77**	P	N				0.57	M	T
Coja26899_c0_seq1	*00000004047	Uncharacterized protein	0.18	2.30	**2.31**	P, GE	N	0.04	0.43	0.09	0.99	U	FG
Coja4779_c0_seq1	*00000003362	Acyl-CoA synthetase short-chain family member 2	0.18	0.77	0.54	P	N	0.02	0.10	0.19	1.00	U	
Coja16408_c0_seq1	*00000002576	Delta-2 crystallin	0.18	0.69	0.60	P	N	0.02	0.15	0.10	0.75	U	FG
Coja12938_c0_seq1	*00000009975	Chaperonin containing TCP1, subunit 2 (beta)	0.18	0.58	0.24	P	N	0.00	0.12	0.01	1.00	U	
Coja14436_c0_seq1	*00000011429	Leukotriene A4 hydrolase	0.18	0.52	0.28	P	N	0.02	0.15	0.14	0.85	U	

Coja15020_c0_seq1	*00000005608	Solute carrier family 9, subfamily A (NHE3, cation proton antiporter 3), member 3 regulator 2	0.18	0.41	**1.12**	P, GE	N	0.01	0.18	0.03	1.00	U	FG
Coja9486_c0_seq1	*00000004852	Uncharacterized protein	0.18	0.28	0.07	P	N				0.91	U	
Coja7443_c0_seq1	*00000015142	Uncharacterized protein	0.18	0.23	0.02	P	Y	0.06	0.13	0.45	1.00	FG	FG
Coja23185_c0_seq1	*00000003495	Aldehyde dehydrogenase 9 family, member A1	0.18	0.21	-0.08	P	Y	0.02	0.17	0.13	0.93	U	L
Coja16668_c0_seq1	*00000005079	Vinculin	0.18	0.19	0.47	P	N	0.00	0.15	0.02	1.00	U	
Coja17098_c0_seq1	*00000008983	Guanine nucleotide binding protein (G protein), beta polypeptide 4	0.18	0.14	0.35	P	N	0.00	0.13	0.00	0.95	U	
Coja263111_c0_seq1	Not ortholog		0.18	0.04	-0.16	P	N				0.33	U	U
Coja24228_c0_seq1	*00000016210	CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase 1	0.17	3.56	**2.02**	P, GE	N	0.03	0.13	0.20	1.00	U	FG
Coja6161_c0_seq1	*00000002103	Tubulin polymerization-promoting protein family member 3	0.17	1.68	-0.63	P	N				0.93	M	FG
Coja5724_c0_seq1	*00000010839	Signal peptidase complex subunit 3 homolog	0.17	1.31	**0.57**	P, GE	N	0.02	0.11	0.17	1.00	U	
Coja43693_c0_seq1	*00000012698*	*Carbohydrate (chondroitin 4) sulfotransferase 11*	0.17	1.25	**6.34**	P, GE	N	0.07	0.39	0.17	0.88	M	FG
Coja15891_c0_seq1	*00000008780	Uncharacterized protein	0.17	0.73	**2.39**	P, GE	Y	0.04	0.16	0.24	0.96	U	M

Coja10963_c0_seq1	Not ortholog		0.17	0.66	-0.10	P	N				0.97	U	U
Coja15301_c0_seq1	*00000011637	T-complex 1	0.17	0.55	0.01	P	N	0.00	0.23	0.02	1.00	U	
Coja42226_c0_seq1	Not ortholog		0.17	0.38	0.57	P	N				0.43	U	
Coja15748_c0_seq2	*00000004874	Peptidyl-prolyl cis-trans isomerase	0.17	0.36	-0.35	P	N	0.00	0.19	0.00	0.99	U	U
Coja13450_c0_seq1	*00000021025	Uncharacterized protein	0.17	0.32	**2.79**	P, GE	N				1.00	FG	FG
Coja9821_c0_seq1	Not ortholog		0.17	0.28	0.33	P	N				0.34	U	
Coja14054_c0_seq1	*00000012245	Alcohol dehydrogenase 5 (class III), chi polypeptide	0.17	0.28	-0.35	P	N	0.03	0.08	0.34	1.00	U	L
Coja48071_c0_seq1	*00000014805	Mak3 homolog	0.17	0.19	-0.04	P	N	0.00	0.02	0.00	0.99	U	
Coja16366_c0_seq1	*00000019436	MAL2 proteolipid protein	0.17	0.19	0.87	P	N	0.03	0.16	0.20	0.99	M	
Coja57721_c0_seq1	*00000015405	Prolyl endopeptidase	0.17	0.18	0.49	P	N	0.01	0.09	0.15	1.00	U	
Coja144873_c0_seq1	*00000015345	Absent in melanoma 1	0.17	0.16	0.55	P	N	0.05	0.14	0.35	1.00	U	FG
Coja3500_c0_seq1	*00000002501	Vitamin K epoxide reductase complex, subunit 1-like 1	0.17	0.13	**0.99**	P, GE	N	0.01	0.14	0.06	1.02	U	
Coja107996_c0_seq1	*00000014496	Prominin 1	0.17	0.13	**4.70**	P, GE	N	0.01	0.09	0.12	0.62	U	
Coja237613_c0_seq1	*00000000750	Calpain 5	0.17	0.09	0.18	P	N	0.01	0.13	0.04	0.80	U	FG
Coja545988_c0_seq1	Not ortholog		0.17	0.02	-0.93	P	N				0.13	U	
Coja11115_c0_seq1	*00000004414	Leucine zipper protein 2	0.17	0.00		P	Y	0.09	0.29	0.32	0.94	T	T
Coja12392_c0_seq3	*00000023986	Uncharacterized protein	0.17	0.00	-0.72	P	Y				0.96	T	T
Coja7905_c0_seq1	*00000004219	Proteasome (prosome, macropain) 26S subunit, non-ATPase, 13	0.16	0.56	-0.10	P	N				1.02	U	

Coja14906_c0_seq1	*00000001862	Alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase 1	0.16	0.53	**2.36**	P, GE	N	0.07	0.10	0.65	1.00	FG	FG
Coja12889_c0_seq1	*00000010298	Dynactin 6	0.16	0.44	0.35	P	N	0.00	0.31	0.01	0.99	U	
Coja4957_c0_seq1	*00000017282	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 9, 39kDa	0.16	0.36	-0.24	P	N	0.03	0.10	0.28	1.01	U	L
Coja24868_c0_seq1	*00000002716	Transmembrane and coiled-coil domains 1	0.16	0.35	0.24	P	Y	0.00	0.21	0.00	1.00	U	
Coja6089_c0_seq1	*00000006061	RAP2C, member of RAS oncogene family	0.16	0.17	0.55	P	N	0.00	0.15	0.03	0.93	U	U
Coja3371_c0_seq1	*00000015473	ADP-ribosylarginine hydrolase	0.16	0.14	-0.26	P	Y	0.05	0.21	0.26	1.00	U	
Coja13205_c0_seq1	*00000016089	Protease-associated domain containing 1	0.16	0.12	0.08	P	Y	0.00	0.30	0.00	0.99	U	
Coja5206_c0_seq1	*00000011798	Heme binding protein 1	0.16	0.11	-0.07	P	N	0.02	0.11	0.19	0.99	U	L
Coja135743_c0_seq1	*00000015177	Guanine nucleotide binding protein (G protein), q polypeptide	0.16	0.08	0.67	P	N	0.00	0.11	0.00	1.00	U	
Coja10373_c0_seq1	No hit	Novel	0.15	8.35	-0.19	P	N					U	
Coja4804_c0_seq1	*00000012172	Ribosomal protein L3	0.15	6.00	-0.09	P	N	0.00	0.19	0.03	1.00	U	
Coja22026_c0_seq1	Not ortholog		0.15	1.48	**1.14**	P, GE	N				0.99	U	
Coja17318_c0_seq1	*00000003835	Sarcoplasmic/endoplasmic reticulum calcium ATPase 2	0.15	0.81	**-2.58**	P	N	0.00	0.12	0.02	1.00	U	
Coja12423_c0_seq1	*00000008985	Sorcin	0.15	0.76	**0.74**	P, GE	N	0.02	0.13	0.16	0.83	U	
Coja10855_c0_seq1	*00000000349	Uncharacterized protein	0.15	0.66	0.16	P	N	0.00	0.23	0.00	1.00	U	
Coja14409_c0_seq1	Not ortholog		0.15	0.66	**1.42**	P, GE	N				1.17	U	FG

Coja16234_c0_seq1	Not ortholog		0.15	0.64	0.37	P	N				0.82	M	FG
Coja5088_c0_seq1	*00000016430	Pyruvate dehydrogenase (lipoamide) alpha 1	0.15	0.56	0.02	P	Y	0.01	0.19	0.06	1.00	U	
Coja10195_c0_seq1	*00000012408	Proteasome (prosome, macropain) 26S subunit, ATPase, 6	0.15	0.48	-0.41	P	N	0.00	0.13	0.00	1.00	U	
Coja28796_c0_seq1	*00000009325	Obg-like ATPase 1	0.15	0.43	**0.59**	P, GE	N	0.00	0.12	0.01	1.00	U	
Coja33366_c0_seq1	**00000013599*	*N-acylsphingosine amidohydrolase (acid ceramidase) 1*	0.15	0.35	-0.18	P	Y	0.02	0.13	0.17	1.00	U	
Coja13785_c0_seq1	*00000003215	Penta-EF-hand domain containing 1	0.15	0.26	0.12	P	N				0.63	U	U
Coja224796_c0_seq1	*00000006722		0.15	0.24	0.27	P	N				0.71	FG	FG
Coja8136_c0_seq1	Not ortholog		0.15	0.22	0.05	P	N				0.97	U	L
Coja5973_c0_seq1	*00000012312	Uncharacterized protein	0.15	0.19	-0.52	P	N	0.01	0.18	0.04	0.90	U	L
Coja125390_c0_seq1	*00000007645	Uncharacterized protein	0.15	0.18	-0.77	P	N				0.97	U	FG
Coja14222_c0_seq1	*00000004707	CSE1 chromosome segregation 1-like (yeast)	0.15	0.17	-0.08	P	N	0.01	0.12	0.08	1.05	U	T
Coja10320_c0_seq1	*00000006179	6S ribosomal protein L13	0.14	8.28	-0.27	P	N	0.01	0.40	0.02	1.00	U	
Coja13764_c0_seq1	No hit	Novel	0.14	4.39	-0.64	P	N					U	
Coja21644_c0_seq1	*00000007089	ER lumen protein retaining receptor	0.14	2.79	**1.44**	P, GE	N	0.00	0.03	0.08	1.00	U	
Coja6158_c0_seq1	*00000016878	Transmembrane 9 superfamily member 2	0.14	2.25	**1.49**	P, GE	Y	0.01	0.12	0.06	0.95	U	FG
Coja6155_c0_seq1	*00000003815	Ribophorin II	0.14	1.98	**1.52**	P, GE	N	0.01	0.11	0.11	0.96	U	
Coja23252_c0_seq1	*00000017675	Uncharacterized protein	0.14	0.78	-0.43	P	N	0.01	0.43	0.03	0.99	U	

Coja12866_c0_seq1	*00000002891	Acyl-Coenzyme A dehydrogenase, long chain	0.14	0.70	**-0.76**	P	N	0.03	0.14	0.19	0.96	U	L
Coja18991_c0_seq1	*00000003475	ATP citrate lyase	0.14	0.66	**1.76**	P, GE	N	0.01	0.19	0.06	1.00	U	L
Coja13999_c0_seq1	*00000005299	Uncharacterized protein	0.14	0.61	0.30	P	N	0.01	0.12	0.12	0.76	U	FG
Coja14874_c0_seq1	*00000000454	Karyopherin (importin) beta 1	0.14	0.46	0.33	P	N	0.00	0.11	0.02	1.00	U	
Coja37195_c0_seq1	*00000012138	Septin 7	0.14	0.41	-0.16	P	N	0.00	0.06	0.02	1.01	U	
Coja83838_c1_seq1	*00000006903	Uncharacterized protein	0.14	0.41	0.39	P	N	0.01	0.22	0.06	0.86	U	FG
Coja44944_c0_seq1	*00000017447	Osteoclast stimulating factor 1	0.14	0.30	0.23	P	N	0.02	0.13	0.12	0.99	U	
Coja12816_c0_seq1	*00000002162	Aconitase 1, soluble	0.14	0.29	**0.67**	P, GE	N	0.01	0.12	0.07	1.00	U	L
Coja11862_c0_seq1	*00000008315	Unc-45 homolog A (C. elegans)	0.14	0.25	0.59	P	N	0.04	0.31	0.12	0.91	U	
Coja16261_c1_seq1	*00000010617	Glutaminyl-peptide cyclotransferase	0.14	0.24	**1.04**	P, GE	N	0.02	0.15	0.14	0.92	U	T
Coja14710_c0_seq1	*00000008701	Xanthine dehydrogenase	0.14	0.21	0.96	P	N	0.01	0.12	0.11	1.00	U	L
Coja19183_c0_seq1	*00000002030	Adenosylhomocysteinase	0.14	0.16	**-0.75**	P	N	0.00	0.12	0.04	0.99	U	L
Coja13807_c0_seq1	*00000012310	V-ral simian leukemia viral oncogene homolog A (ras related)	0.14	0.13	-0.06	P	N	0.00	0.09	0.00	1.00	U	
Coja4907_c0_seq1	*00000022586	Hemopexin	0.14	0.08	-0.85	P	Y	0.06	0.42	0.15	0.98	U	L
Coja464905_c0_seq1	Not ortholog		0.14	0.05	-0.89	P	N				0.75	M	FG
Coja354120_c0_seq1	No hit	Novel	0.14	0.03	**-1.44**	P	N					U	
Coja328030_c0_seq1	*00000009444	HHIP-like 2	0.14	0.01	-0.27	P	N	0.14	0.33	0.43	0.23	M	
Coja4906_c0_seq1	No hit	Novel	0.13	8.25	-0.21	P	N					U	
Coja18759_c0_seq1	*00000008590	Eukaryotic translation elongation factor 1 beta 2	0.13	5.40	-0.34	P	N	0.00	0.12	0.00	1.00	U	
Coja17340_c0_seq1	*00000005966	Uncharacterized protein	0.13	1.41	**1.77**	P, GE	N	0.00	0.15	0.03	0.99	U	FG

Coja21646_c0_seq1	*00000022570	Hepatoma-derived growth factor	0.13	1.26	-0.20	P	N	0.01	0.15	0.10	0.87	U	
Coja12142_c0_seq1	*00000002750	NCK-associated protein 1	0.13	0.63	**1.21**	P, GE	N	0.00	0.18	0.01	1.00	U	
Coja14368_c0_seq1	**00000002914*	*Vesicle-trafficking protein SEC22b*	0.13	0.55	**1.23**	P, GE	Y	0.00	0.36	0.00	1.00	U	
Coja15726_c0_seq1	*00000007330	Uncharacterized protein	0.13	0.54	0.21	P	N	0.01	0.13	0.04	1.00	U	
Coja20964_c0_seq1	*00000011661	Superoxide dismutase 2, mitochondrial	0.13	0.50	-0.58	P	N	0.01	0.14	0.06	1.00	U	L
Coja5539_c0_seq1	*00000000065	Proteasome alpha 5 subunit	0.13	0.42	**-0.65**	P	N	0.01	0.23	0.03	1.00	U	
Coja10525_c0_seq2	**00000012233*	*Uncharacterized protein*	0.13	0.37	0.30	P	N	0.01	0.04	0.30	0.98	U	T
Coja10921_c0_seq1	*00000007664	Proteasome (prosome, macropain) 26S subunit, non-ATPase, 1	0.13	0.37	0.23	P	N	0.00	0.12	0.03	1.00	U	U
Coja4627_c0_seq1	Not ortholog		0.13	0.37	-0.04	P	N				0.93	U	U
Coja5976_c0_seq1	*00000012404	ERO1-like (S. cerevisiae)	0.13	0.36	0.83	P	Y	0.03	0.20	0.14	0.95	U	
Coja16500_c0_seq1	*00000015664	Adducin 1 (alpha)	0.13	0.34	0.19	P	N	0.02	0.11	0.17	0.92	U	
Coja14342_c0_seq1	*00000006799	Adaptor-related protein complex 2, alpha 2 subunit	0.13	0.28	0.31	P	N	0.00	0.07	0.01	1.00	U	
Coja98682_c0_seq1	*00000017205	Endonuclease domain containing 1	0.13	0.21	0.21	P	Y	0.03	0.08	0.34	1.00	M	FG
Coja15715_c0_seq1	*00000003281	Glutathione synthetase	0.13	0.17	-0.20	P	N	0.02	0.16	0.11	1.00	U	
Coja18243_c0_seq1	Not ortholog		0.13	0.15	**-1.42**	P	N				0.99	U	L
Coja7917_c0_seq1	**00000005097*	*Uncharacterized protein*	0.13	0.05	-1.39	P	N				0.97	U	L
Coja4849_c0_seq1	*00000004591	Antithrombin	0.13	0.03	-0.51	P	Y	0.03	0.20	0.18	0.95	U	L
Coja370572_c0_seq1	Not ortholog		0.13	0.00		P	N				0.55	T	T
Coja17905_c0_seq1	*00000015082	Ribosomal protein S6	0.12	16.99	-0.12	P	N	0.00	0.14	0.01	0.91	U	

Coja19902_c0_seq1	*00000015003	ATPase, Na+/K+ transporting, alpha 1 polypeptide	0.12	4.32	**1.34**	P, GE	N	0.00	0.08	0.04	1.00	U	
Coja19651_c0_seq1	*00000008229	PREDICTED: similar to Antiquitin	0.12	2.03	**1.02**	P, GE	Y	0.02	0.12	0.14	1.00	U	L
Coja26958_c0_seq1	*00000007430	Coatomer subunit delta	0.12	1.46	**1.43**	P, GE	N	0.00	0.11	0.01	1.00	U	
Coja4692_c0_seq1	*00000000905	Proteasome (prosome, macropain) subunit, beta type, 4	0.12	0.89	-0.19	P	N	0.03	0.24	0.14	1.00	U	
Coja14161_c0_seq1	*00000010547	Protein YIPF4	0.12	0.84	**1.32**	P, GE	N	0.00	0.12	0.03	0.86	U	
Coja28456_c0_seq1	Not ortholog		0.12	0.63	-0.62	P	N				0.98	U	
Coja11913_c0_seq1	*00000023510	Uncharacterized protein	0.12	0.46	0.56	P	Y				0.88	U	
Coja10316_c0_seq1	*00000010652	Uncharacterized protein	0.12	0.42	0.44	P	N	0.01	0.22	0.05	0.98	U	L
Coja15622_c0_seq1	Not ortholog		0.12	0.37	0.25	P	Y				0.43	U	
Coja13220_c0_seq1	*00000001811	Integrin beta 4 binding protein	0.12	0.30	0.25	P	N	0.00	0.17	0.03	1.00	U	
Coja14337_c0_seq1	*00000009112	Leucine rich repeat containing 57	0.12	0.30	**1.29**	P, GE	N	0.01	0.09	0.16	1.00	U	T
Coja6187_c0_seq1	*00000007543	Uncharacterized protein	0.12	0.30	1.08	P	Y	0.02	0.25	0.07	0.96	M	M
Coja13914_c0_seq1	*00000011704	Protein tyrosine phosphatase-like (proline instead of catalytic arginine), member b	0.12	0.28	**1.15**	P, GE	N	0.01	0.13	0.09	0.96	U	
Coja15338_c0_seq1	Not ortholog		0.12	0.25	-0.45	P	N				0.24	U	
Coja16767_c0_seq1	*00000005147	Amylase, alpha 1A salivary	0.12	0.24	0.20	P	Y	0.05	0.33	0.15	1.00	U	
Coja25723_c0_seq1	*00000017366	Dolichyl-phosphate mannosyltransferase polypeptide 1, catalytic subunit	0.12	0.23	-0.34	P	N	0.10	0.38	0.26	0.85	U	
Coja3282_c0_seq1	*00000010053		0.12	0.22	0.54	P	Y				0.83	U	FG

Coja9242_c0_seq1	*00000010150	Sec23 homolog A (S. cerevisiae)	0.12	0.20	0.01	P	N	0.00	0.13	0.00	1.00	U	
Coja12113_c0_seq1	*00000004730	Chloride intracellular channel 2	0.12	0.13	**-1.14**	P	N	0.00	0.38	0.00	1.00	U	L
Coja5532_c0_seq1	**00000003178*	*Surfeit locus protein 4*	0.11	1.42	**1.75**	P, GE	N	0.00	0.06	0.00	1.00	U	
Coja15380_c0_seq1	*00000001331	PREDICTED: similar to epithin	0.11	1.18	**1.47**	P, GE	N	0.03	0.21	0.15	0.98	U	FG
Coja16316_c0_seq1	*00000014933	Hexosaminidase B (beta polypeptide)	0.11	1.08	**1.43**	P, GE	Y	0.06	0.18	0.31	1.00	U	
Coja44500_c0_seq1	*00000005318	Uncharacterized protein	0.11	0.72	**4.08**	P, GE	Y	0.04	0.13	0.27	1.01	M	FG
Coja24237_c0_seq1	*00000001654	Complement component 1, q subcomponent binding protein	0.11	0.69	-0.38	P	N	0.01	0.36	0.03	0.99	U	
Coja12043_c0_seq1	Not ortholog		0.11	0.69	0.23	P	N				0.93	U	T
Coja88486_c0_seq1	*00000002452	Uncharacterized protein	0.11	0.64	-0.48	P	N	0.01	0.16	0.04	1.00	M	FG
Coja24293_c0_seq1	*00000010671	Uncharacterized protein	0.11	0.56	-0.21	P	N	0.10	0.19	0.52	0.54	U	
Coja15163_c0_seq1	*00000006427	Cysteinyl-tRNA synthetase, cytoplasmic	0.11	0.41	**1.30**	P, GE	N	0.00	0.10	0.04	1.00	U	
Coja46232_c0_seq1	*00000000814	Heterogeneous nuclear ribonucleoprotein R	0.11	0.40	0.12	P	N	0.00	0.05	0.03	0.69	U	
Coja4966_c0_seq1	*00000004377	Uncharacterized protein	0.11	0.37	0.10	P	N	0.00	0.12	0.00	1.00	U	
Coja12967_c0_seq1	**00000007163*	*Platelet-activating factor acetylhydrolase, isoform Ib, beta subunit 3kDa*	0.11	0.34	**0.64**	P, GE	N	0.01	0.04	0.21	1.00	U	T
Coja8375_c0_seq1	*00000003563	Tyrosyl-tRNA synthetase	0.11	0.34	**0.83**	P, GE	N	0.01	0.11	0.10	1.00	U	

Coja17539_c0_seq2	*00000006598	Sortilin-related receptor, L(DLR class) A repeats containing	0.11	0.33	**1.68**	P, GE	N	0.01	0.19	0.04	1.00	M	FG
Coja8192_c0_seq1	No hit	Novel	0.11	0.33	-0.29	P	N					U	
Coja13230_c0_seq1	*00000000259	Proteasome inhibitor subunit 1	0.11	0.25	-0.18	P	N	0.03	0.13	0.27	1.00	U	
Coja14123_c0_seq1	*00000014525	Ubiquitin carboxyl-terminal hydrolase	0.11	0.23	-0.04	P	N	0.00	0.12	0.02	1.00	U	
Coja11868_c0_seq1	*00000003316	Flavin containing monooxygenase 6	0.11	0.22	-0.68	P	N	0.04	0.22	0.16	0.96	U	L
Coja162113_c0_seq1	Not ortholog		0.11	0.19	**0.86**	P, GE	N				0.67	M	FG
Coja13349_c0_seq1	*00000013863	Proteasome (prosome, macropain) assembly chaperone 2	0.11	0.18	0.00	P	N	0.03	0.15	0.23	1.03	U	
Coja15175_c0_seq1	*00000023768	Uncharacterized protein	0.11	0.17	-0.29	P	N				1.00	U	
Coja265286_c0_seq1	*00000023421	CUB and zona pellucida-like domains 1	0.11	0.13	**1.98**	P, GE	N	0.09	0.25	0.37	0.97	FG	FG
Coja12806_c0_seq1	Not ortholog		0.11	0.12	-0.31	P	N				0.96	U	L
Coja15970_c0_seq1	*00000005321	Laminin, alpha 5	0.11	0.11	0.21	P	N	0.03	0.15	0.18	0.44	U	
Coja258495_c0_seq1	*00000016682	Cysteine-rich secretory protein 1	0.11	0.07		P	Y						
Coja17706_c0_seq1	*00000009118	PIT 54 protein	0.11	0.03	-0.63	P	N	0.08	0.27	0.28	0.98	U	L
Coja10345_c0_seq1	*00000016131	Uncharacterized protein	0.10	5.66	-0.44	P	N	0.02	0.27	0.09	1.01	U	
Coja10159_c0_seq1	*00000003153	Coatomer subunit epsilon	0.10	1.29	**0.98**	P, GE	N	0.01	0.26	0.05	0.94	U	
Coja12300_c1_seq1	*00000023199	Heterogeneous nuclear ribonucleoprotein D-like	0.10	1.15	0.00	P	N	0.00	0.22	0.00	1.00	U	
Coja11653_c0_seq1	*00000001413	*Meteorin, glial cell differentiation regulator-like*	0.10	0.68	0.49	P	Y	0.02	0.08	0.26	0.98	U	FG

Coja20085_c0_seq1	*00000001240	Uncharacterized protein	0.10	0.55	**0.83**	P, GE	N	0.00	0.20	0.02	0.97	U	L
Coja12399_c0_seq1	*00000002388	Hypothetical protein LOC417539	0.10	0.53	-0.47	P	N	0.01	0.04	0.28	1.00	U	
Coja11973_c0_seq1	*00000004817	Gem (nuclear organelle) associated protein 4	0.10	0.46	0.29	P	N	0.05	0.19	0.27	1.00	U	
Coja12561_c0_seq1	Not ortholog		0.10	0.44	-0.20	P	N				0.46	U	
Coja12904_c0_seq1	*00000015348	Activated leukocyte cell adhesion molecule	0.10	0.42	-0.17	P	Y	0.01	0.09	0.11	0.92	U	FG
Coja27198_c0_seq1	*00000000522		0.10	0.42	-0.25	P	N				0.97	U	
Coja15223_c0_seq1	*00000008523	Uncharacterized protein	0.10	0.41	**0.74**	P, GE	N	0.01	0.10	0.07	1.00	U	
Coja14481_c0_seq1	No hit	Novel	0.10	0.36	0.61	P	N					U	FG
Coja3042_c0_seq1	Not ortholog		0.10	0.30	0.09	P	N				0.69	U	FG
Coja35309_c0_seq1	Not ortholog		0.10	0.28	**0.66**	P, GE	N				0.96	U	T
Coja9720_c0_seq1	*00000006805	Phospholipid scramblase 4	0.10	0.28	-0.26	P	N	0.03	0.08	0.30	1.00	U	L
Coja6797_c0_seq1	Not ortholog		0.10	0.21	-0.27	P	N				0.27	U	FG
Coja12758_c0_seq1	*00000002484	Crystallin, mu	0.10	0.16	0.82	P	N				1.00	U	
Coja8453_c0_seq1	*00000004882	Phosphopantothenoylcysteine synthetase	0.10	0.14	**0.56**	P, GE	N	0.04	0.46	0.08	0.95	U	
Coja15604_c0_seq1	*00000006839	Diaphanous 1	0.10	0.12	0.32	P	N	0.06	0.38	0.17	0.69	U	T
Coja13951_c0_seq2	*00000020488		0.10	0.08	**0.77**	P, GE	Y				0.67	U	
Coja12001_c0_seq1	*00000014372	Solute carrier family 34 (sodium phosphate), member 2	0.09	1.82	**3.51**	P, GE	N	0.01	0.09	0.12	1.00	FG	FG
Coja10288_c0_seq2	*00000011036	Heterogeneous nuclear ribonucleoprotein A2/B1	0.09	1.62	-0.13	P	N	0.00	0.09	0.00	0.55	U	
Coja5918_c0_seq1	*00000007705	Nucleolin	0.09	1.27	0.17	P	N	0.03	0.14	0.19	0.65	U	

Coja14424_c0_seq1	*00000010744	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 7 (GalNAc-T7)	0.09	1.20	**1.65**	P, GE	N	0.02	0.18	0.10	1.00	M	FG
Coja20712_c0_seq1	*00000002363	Heat shock 7kDa protein 9B	0.09	1.10	0.00	P	N	0.01	0.12	0.04	0.96	U	
Coja21591_c0_seq1	*00000001364	Casein kinase I isoform alpha	0.09	1.08	**0.70**	P, GE	N	0.00	0.05	0.00	0.98	U	
Coja13335_c0_seq1	*00000006728	PREDICTED: similar to sodium-glucose cotransporter-like 1	0.09	0.91	**1.52**	P, GE	N	0.02	0.09	0.22	0.98	M	FG
Coja15012_c0_seq1	*00000010878	Similar to Vesicle docking protein p115	0.09	0.89	**0.99**	P, GE	N	0.01	0.13	0.06	1.00	U	
Coja14657_c0_seq1	Not ortholog		0.09	0.56	**1.25**	P, GE	Y				1.01	U	
Coja17198_c0_seq1	*00000004814	Rhophilin, Rho GTPase binding protein 2	0.09	0.51	**1.74**	P, GE	N	0.03	0.15	0.18	0.99	U	
Coja22115_c0_seq1	*00000012610	Uncharacterized protein	0.09	0.44	-0.05	P	Y	0.01	0.10	0.09	0.95	U	L
Coja22469_c0_seq1	*00000010463	AHA1, activator of heat shock 9kDa protein ATPase homolog 1 (yeast)	0.09	0.43	-0.54	P	N	0.01	0.13	0.11	1.00	U	
Coja14146_c0_seq1	*00000009551	Dynein, cytoplasmic, intermediate polypeptide 2	0.09	0.40	**0.74**	P, GE	N	0.00	0.12	0.03	1.03	U	
Coja61423_c0_seq1	*00000002022	UDP-Gal:betaGlcNAc beta 1,4-galactosyltransferase, polypeptide 1	0.09	0.26	0.66	P	N	0.03	0.18	0.16	0.98	U	
Coja17043_c0_seq1	*00000001615	Mannosidase, alpha, class 2C, member 1	0.09	0.15	0.41	P	N	0.02	0.15	0.14	1.00	U	

Coja15023_c0_seq1	*00000010593	Galactosylceramidase	0.09	0.14	**1.84**	P, GE	N	0.04	0.24	0.18	1.00	M	T
Coja4994_c0_seq1	*00000009728	Solute carrier family 25, member 13 (citrin)	0.09	0.13	**0.76**	P, GE	N	0.01	0.08	0.07	0.99	U	L
Coja6780_c0_seq1	*00000001832	Phytanoyl-CoA 2-hydroxylase interacting protein-like	0.09	0.13	0.04	P	N	0.00	0.05	0.04	0.88	M	M
Coja71660_c0_seq1	No hit	Novel	0.09	0.12	-0.62	P	N					U	U
Coja17235_c0_seq2	*00000001558	Mov1, Moloney leukemia virus 1, homolog	0.09	0.12	-0.01	P	N	0.04	0.31	0.14	1.00	U	L
Coja10395_c0_seq1	*00000016885	Serine/threonine kinase 24 (STE2 homolog, yeast)	0.09	0.08	0.59	P	N	0.01	0.12	0.07	0.79	M	
Coja17799_c0_seq1	*00000012613	Uncharacterized protein	0.09	0.08	-0.52	P	N	0.00	0.14	0.03	0.97	U	L
Coja449613_c0_seq1	Not ortholog		0.09	0.06	**-2.14**	P	N				0.19	U	
Coja20467_c0_seq1	*00000005955	Heterogeneous nuclear ribonucleoprotein H1 (H)	0.08	1.80	0.24	P	N	0.02	0.21	0.09	0.69	U	
Coja27025_c0_seq1	*00000001005	Uncharacterized protein	0.08	1.51	**1.23**	P, GE	N	0.00	0.28	0.00	1.00	U	FG
Coja6033_c0_seq1	*00000002500	GDP-mannose pyrophosphorylase B	0.08	1.43	**1.44**	P, GE	N	0.00	0.33	0.01	1.00	U	FG
Coja4685_c0_seq1	*00000005807	Importin 7	0.08	0.69	0.26	P	N	0.00	0.10	0.01	1.00	U	
Coja13870_c0_seq1	*00000005694	Glycyl-tRNA synthetase	0.08	0.58	0.73	P	Y	0.01	0.15	0.08	0.92	U	
Coja10260_c0_seq1	*00000005057	Adenosine kinase	0.08	0.54	**0.64**	P, GE	N	0.01	0.08	0.08	1.00	U	L
Coja6792_c0_seq1	*00000014395	DEAH (Asp-Glu-Ala-His) box polypeptide 15	0.08	0.43	**0.77**	P, GE	N	0.00	0.08	0.01	1.00	U	
Coja18025_c0_seq1	*00000003678	Glutamine synthetase	0.08	0.41	-0.49	P	N	0.00	0.12	0.02	1.00	U	L

Coja16618_c0_seq1	*00000002310	3-hydroxyisobutyryl-Coenzyme A hydrolase	0.08	0.40	0.23	P	N	0.04	0.13	0.33	1.00	U	L
Coja15665_c0_seq1	*00000009953	Leucine rich repeat containing 19	0.08	0.40	**1.04**	P, GE	N	0.05	0.15	0.35	1.06	U	L
Coja14944_c0_seq1	*00000008606	NADH dehydrogenase (ubiquinone) Fe-S protein 1, 75kDa (NADH-coenzyme Q reductase)	0.08	0.39	-0.33	P	N	0.01	0.16	0.06	0.98	U	
Coja13001_c0_seq1	*00000011570	Interleukin enhancer binding factor 2, 45kDa	0.08	0.35	-0.32	P	N	0.01	0.23	0.02	0.84	U	
Coja4511_c0_seq1	*00000003303	Uncharacterized protein	0.08	0.34	-0.17	P	N	0.02	0.16	0.14	0.98	U	
Coja16497_c0_seq1	Not ortholog		0.08	0.31	0.25	P	N				1.00	U	
Coja15485_c0_seq1	*00000016887	Uncharacterized protein	0.08	0.30	**-1.45**	P	N				0.99	U	T
Coja15063_c0_seq2	*00000018747	Hydroxysteroid (17-beta) dehydrogenase 1	0.08	0.21	-0.11	P	N				0.98	U	
Coja146095_c0_seq1	*00000005065	Cytosolic phospholipase A2Phospholipase A2Lysophospholipase	0.08	0.20	**0.93**	P, GE	N	0.01	0.15	0.04	1.00	U	FG
Coja16315_c0_seq1	*00000013091	Epidermal growth factor receptor pathway substrate 8	0.08	0.18	**0.71**	P, GE	N	0.00	0.10	0.04	1.00	U	FG
Coja13773_c0_seq1	*00000001521	Vacuolar protein sorting 26 homolog B (S. pombe)	0.08	0.12	0.59	P	N	0.00	0.26	0.01	0.98	U	

Coja251138_c0_seq1	*00000016412	Membrane-bound O-acyltransferase domain-containing protein 2	0.08	0.09	0.10	P	N	0.01	0.08	0.10	0.67	M	FG
Coja10617_c0_seq1	*00000016351	Acyl-Coenzyme A thioesterase 2, mitochondrial	0.08	0.06	-0.39	P	N	0.01	0.09	0.08	0.92	U	
Coja284705_c0_seq1	Not ortholog		0.08	0.06	0.47	P	N				0.20	U	FG
Coja14603_c0_seq1	Not ortholog		0.08	0.04	0.19	P	N				0.41	M	
Coja297_c0_seq1	*00000007068*	*Chromobox homolog 8*	0.08	0.03	-0.10	P	N	0.01	0.19	0.06	1.00	U	
Coja15817_c0_seq1	*00000000293	Uncharacterized protein	0.08	0.02	1.07	P	Y				1.01	M	L
Coja13622_c0_seq1	*00000016447	Uncharacterized protein	0.07	2.93	**1.38**	P, GE	Y	0.01	0.15	0.06	0.95	U	
Coja16084_c0_seq1	*00000005550	ST6 beta-galactosamide alpha-2,6-sialyltranferase 1	0.07	2.91	0.53	P	N	0.04	0.15	0.24	1.00	U	FG
Coja10293_c0_seq1	*00000010641	Saccharopine dehydrogenase (putative)	0.07	1.97	**1.92**	P, GE	N	0.01	0.12	0.06	1.00	U	
Coja45227_c0_seq1	*00000023689	Argininosuccinate synthase	0.07	1.39	**2.00**	P, GE	N	0.03	0.24	0.14	1.00	U	FG
Coja15121_c0_seq1	*00000015904	Transmembrane protein 3A	0.07	0.74	**0.96**	P, GE	N	0.01	0.15	0.05	1.00	U	
Coja10837_c0_seq1	*00000010809	Basic leucine zipper and W2 domains 2	0.07	0.55	-0.13	P	N	0.00	0.04	0.00	0.97	U	T
Coja10870_c0_seq1	*00000003522	Retinoblastoma binding protein 4	0.07	0.49	-0.25	P	N	0.00	0.15	0.00	1.00	U	
Coja15847_c0_seq1	*00000009015	Phosphatidylinositol glycan anchor biosynthesis, class K	0.07	0.40	**0.77**	P, GE	Y	0.02	0.13	0.13	0.95	U	
Coja10503_c0_seq1	*00000006197	Casein kinase II subunit alpha	0.07	0.34	0.15	P	N	0.00	0.04	0.00	1.00	U	U

Coja16321_c0_seq1	*00000000848	Adaptor-related protein complex 1, gamma 1 subunit	0.07	0.32	**0.55**	P, GE	N	0.00	0.07	0.07	1.00	U	
Coja4631_c0_seq1	*00000002531	Uncharacterized protein	0.07	0.31	0.06	P	N	0.00	0.13	0.00	1.00	U	U
Coja14528_c0_seq1	*00000000785	Hyaluronoglucosaminidase 1	0.07	0.26	0.14	P	Y				0.92	U	
Coja12111_c0_seq1	*00000001329	Erythroid protein 4.1	0.07	0.25	0.40	P	N	0.01	0.14	0.06	1.00	U	
Coja5077_c0_seq1	*00000000142	Choline/ethanolamine phosphotransferase 1	0.07	0.25	0.50	P	N	0.03	0.16	0.19	1.00	U	L
Coja14328_c0_seq1	*00000012137	Kinectin	0.07	0.24	-0.30	P	N	0.02	0.08	0.25	1.00	U	
Coja14147_c0_seq1	*00000012365	Liver glycogen phosphorylase	0.07	0.20	0.41	P	N	0.01	0.24	0.05	1.00	U	L
Coja15506_c0_seq1	*00000001700	Uncharacterized protein	0.07	0.20	-0.19	P	N	0.02	0.27	0.09	1.00	U	
Coja4162_c0_seq1	*00000008930	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 2	0.07	0.19	0.55	P	N	0.01	0.14	0.10	1.00	U	
Coja13863_c0_seq1	*00000017011	Tripartite motif containing 13	0.07	0.18	0.35	P	N	0.01	0.13	0.08	0.97	U	
Coja31276_c0_seq1	*00000008144	Coatomer subunit gamma	0.07	0.18	0.20	P	N	0.00	0.15	0.00	1.00	U	T
Coja4518_c0_seq1	*00000008559	Uncharacterized protein	0.07	0.17	**0.69**	P, GE	N	0.00	0.09	0.01	1.00	U	T
Coja15071_c0_seq4	*00000002784	N-acylaminoacyl-peptide hydrolase	0.07	0.14	-0.05	P	N				0.98	U	
Coja15905_c0_seq1	*00000009536	Transcobalamin II	0.07	0.12	-0.33	P	Y	0.06	0.23	0.25	1.00	U	L
Coja11778_c0_seq1	*00000004109*	*Guanine nucleotide binding protein (G protein), alpha 13*	0.07	0.11	**0.80**	P, GE	N	0.00	0.22	0.01	1.00	U	
Coja12599_c0_seq1	*00000006345	Transmembrane protein 43	0.07	0.09	**-0.69**	P	N	0.01	0.09	0.17	1.00	U	U
Coja217558_c0_seq1	*00000009885	Helicase (DNA) B	0.07	0.02	-0.49	P	N	0.03	0.14	0.22	0.34	M	U

Coja13961_c0_seq1	*00000016833	Coagulation factor VII precursor	0.07	0.01		P	Y	0.03	0.11	0.31	1.00	L	L
Coja110409_c0_seq1	*00000010057	Uncharacterized protein	0.07	0.00		P	N	0.13	0.30	0.44	0.89	T	T
Coja21377_c0_seq1	*00000016093	Eukaryotic translation initiation factor 3, subunit 6 48kDa	0.06	1.33	-0.37	P	N	0.00	0.13	0.00	1.00	U	
Coja16567_c0_seq1	*00000008469	Serine peptidase inhibitor, Kunitz type 1	0.06	1.18	**1.35**	P, GE	Y	0.03	0.13	0.21	0.99	U	FG
Coja4772_c0_seq1	*00000008812	Chaperonin containing TCP1, subunit 4 (delta)	0.06	0.86	0.21	P	N	0.00	0.19	0.02	0.99	U	
Coja10104_c0_seq1	*00000016340	Eukaryotic translation initiation factor 2, subunit 3 gamma, 52kDa	0.06	0.81	-0.25	P	N	0.00	0.10	0.00	1.00	U	
Coja39985_c0_seq1	*00000002664	Annexin A11	0.06	0.69	-0.08	P	N	0.02	0.15	0.11	0.99	U	
Coja17328_c1_seq1	*00000005469*	*Uncharacterized protein*	0.06	0.64	**1.47**	P, GE	N	0.23	0.48	0.47	0.98	U	L
Coja16561_c0_seq1	*00000006915	BAI1-associated protein 2	0.06	0.52	**0.79**	P, GE	N	0.04	0.20	0.21	0.94	U	FG
Coja9326_c0_seq1	*00000004259	Bleomycin hydrolase	0.06	0.41	**0.62**	P, GE	N	0.01	0.15	0.09	1.00	U	
Coja33816_c0_seq1	*00000022293	Proteasome (prosome, macropain) 26S subunit, non-ATPase, 3	0.06	0.41	-0.35	P	N				0.98	U	
Coja14485_c0_seq1	*00000000674	Myosin ID	0.06	0.40	0.53	P	N	0.02	0.28	0.07	0.98	U	
Coja9034_c0_seq1	*00000010471	Protein phosphatase 2, regulatory subunit B, delta isoform	0.06	0.34	0.32	P	N	0.01	0.25	0.04	0.97	U	
Coja138361_c0_seq1	*00000006635	Rhesus blood group, C glycoprotein	0.06	0.33	0.06	P	N	0.01	0.15	0.05	1.00	M	FG

Coja5146_c0_seq1	*00000000404	COP9 constitutive photomorphogenic homolog subunit 2	0.06	0.29	-0.11	P	N	0.00	0.05	0.00	0.94	U	
Coja13530_c0_seq1	*00000015833	Uncharacterized protein	0.06	0.26	0.86	P	N	0.02	0.16	0.15	0.79	M	FG
Coja3060_c0_seq1	Not ortholog		0.06	0.24	0.37	P	N				1.03	U	
Coja11909_c0_seq1	Not ortholog		0.06	0.24	**-0.59**	P	Y				0.80	U	L
Coja15778_c0_seq1	Not ortholog		0.06	0.21	0.31	P	N				0.36	U	
Coja15962_c0_seq1	*00000001717	Solute carrier family 16, member 1	0.06	0.20	0.61	P	N	0.05	0.18	0.25	1.00	U	T
Coja15162_c0_seq1	*00000002140	Adaptor-related protein complex 2, beta 1 subunit	0.06	0.20	0.40	P	N	0.00	0.11	0.00	1.00	U	T
Coja3360_c0_seq1	*00000011329	Myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila); translocated to, 4	0.06	0.19	0.60	P	N	0.01	0.13	0.11	0.76	U	
Coja6325_c0_seq1	*00000007904	Uncharacterized protein	0.06	0.19	-0.39	P	N	0.01	0.12	0.11	0.87	U	
Coja3877_c0_seq1	*00000019201	Stress 7 protein chaperone, microsome-associated, 6kDa	0.06	0.17	**2.14**	P, GE	Y	0.01	0.14	0.09	1.00	U	
Coja285369_c0_seq1	*00000015140	Uncharacterized protein	0.06	0.16	0.37	P	N	0.03	0.15	0.23	0.54	M	FG
Coja132859_c0_seq1	*00000014846	3-oxoacid CoA transferase 1	0.06	0.15	-0.59	P	N	0.01	0.13	0.07	1.00	M	FG
Coja17115_c0_seq1	*00000012003	Uncharacterized protein	0.06	0.14	-0.32	P	N	0.02	0.10	0.23	0.99	U	
Coja12799_c0_seq1	*00000004106	7-dehydrocholesterol reductase	0.06	0.13	0.41	P	N	0.02	0.20	0.11	1.00	U	L
Coja10429_c0_seq1	Not ortholog		0.06	0.11	0.47	P	Y				0.70	U	

Coja17273_c0_seq1	*00000014716	Interleukin 6 signal transducer (gp13, oncostatin M receptor)	0.06	0.10	0.26	P	Y	0.03	0.13	0.26	1.00	U	
Coja14032_c0_seq1	*00000004384	Solute carrier family 19 (folate transporter), member 1	0.06	0.09	-0.04	P	N	0.04	0.26	0.15	1.01	U	L
Coja10882_c0_seq1	*00000007334	Uncharacterized protein	0.06	0.03	0.01	P	N	0.01	0.13	0.04	1.00	M	L
Coja200136_c0_seq1	Not ortholog		0.06	0.01	-0.03	P	N						
Coja16718_c0_seq1	*00000012108	Glutamyl aminopeptidase (aminopeptidase A)	0.06	0.01	0.42	P	N	0.04	0.11	0.33	1.00	L	L
Coja13948_c0_seq1	*00000010469	Uncharacterized protein	0.06	0.00	-0.80	P	N	0.03	0.11	0.31	0.99	M	L
Coja16601_c0_seq1	*00000006665	Polypeptide N-acetylgalactosaminyltransferase 6	0.05	0.71	**2.02**	P, GE	N	0.01	0.30	0.02	0.89	FG	FG
Coja10082_c0_seq1	*00000008094	6 kDa heat shock protein, mitochondrial	0.05	0.66	-0.03	P	N	0.00	0.19	0.01	0.94	U	L
Coja12702_c0_seq1	*00000009961	ATP-binding cassette, sub-family E, member 1	0.05	0.63	**0.94**	P, GE	N	0.00	0.13	0.00	0.98	U	
Coja5181_c0_seq1	*00000007307	Ubiquitin specific protease 7	0.05	0.55	0.34	P	N	0.00	0.09	0.01	0.99	U	
Coja45418_c0_seq1	*00000007494	Uncharacterized protein	0.05	0.47	**1.04**	P, GE	N	0.01	0.14	0.10	1.00	U	
Coja11081_c0_seq1	*00000005202	Damage-specific DNA binding protein 1, 127kDa	0.05	0.40	0.00	P	N	0.00	0.18	0.01	1.00	U	
Coja18657_c0_seq1	*00000016285	Uncharacterized protein	0.05	0.29	**1.28**	P, GE	N	0.02	0.12	0.18	1.00	U	T
Coja14990_c0_seq1	Not ortholog		0.05	0.25	-0.18	P	N				0.13	M	FG
Coja15727_c0_seq1	*00000016224	Monoamine oxidase A	0.05	0.19	-0.23	P	N	0.02	0.08	0.26	0.99	U	L

Coja139245_c0_seq1	*00000013875	Family with sequence similarity 83, member H	0.05	0.17	0.36	P	N	0.01	0.26	0.06	1.00	U	FG
Coja14688_c1_seq1	Not ortholog		0.05	0.15	-0.33	P	N				0.47	U	U
Coja9012_c0_seq1	Not ortholog		0.05	0.15	0.34	P	N				0.47	U	
Coja14376_c0_seq1	*00000002444	Myosin XVB pseudogene	0.05	0.13	0.44	P	N	0.04	0.12	0.35	1.01	FG	FG
Coja14619_c0_seq1	*00000000374	Arginyl aminopeptidase (aminopeptidase B)	0.05	0.11	**-0.65**	P	N	0.02	0.31	0.08	0.98	U	L
Coja16018_c0_seq1	*00000009082	AarF domain containing kinase 3	0.05	0.10	**-2.31**	P	N	0.01	0.12	0.06	0.99	U	
Coja163868_c0_seq1	*00000006439	Rac/Cdc42 guanine nucleotide exchange factor (GEF) 6	0.05	0.08	**-0.85**	P	N	0.00	0.14	0.03	0.80	U	
Coja6924_c0_seq1	*00000006140	Golgi-associated, gamma adaptin ear containing, ARF binding protein 2	0.05	0.03	**-1.53**	P	N	0.08	0.18	0.42	0.57	U	
Coja10390_c0_seq1	*00000012126	Uncharacterized protein	0.05	0.01	**-0.97**	P	Y	0.03	0.10	0.29	1.00	M	L
Coja14351_c0_seq1	Not ortholog		0.05	0.01	-0.97	P	Y				0.95	M	L
Coja16875_c0_seq1	*00000004295	Carboxypeptidase D	0.04	1.37	**1.84**	P, GE	Y	0.02	0.15	0.11	0.98	U	FG
Coja10204_c0_seq1	*00000016142	Mx protein	0.04	0.92	1.20	P	N	0.15	0.12	1.25	0.93	U	L
Coja17430_c1_seq1	*00000004269	Uncharacterized protein	0.04	0.74	0.17	P	N	0.06	0.18	0.32	0.84	U	U
Coja10434_c0_seq1	*00000013167	Succinate dehydrogenase complex, subunit A, flavoprotein (Fp)	0.04	0.59	-0.56	P	N	0.01	0.15	0.05	1.00	U	
Coja9384_c0_seq1	*00000009275	Transmembrane protein 63A	0.04	0.58	**1.42**	P, GE	N	0.04	0.11	0.33	1.00	U	FG
Coja4724_c0_seq1	*00000005108	Uncharacterized protein	0.04	0.57	0.20	P	N	0.00	0.05	0.00	1.00	U	T
Coja14783_c0_seq1	*00000000357	Uncharacterized protein	0.04	0.35	**2.82**	P, GE	Y				0.96	FG	FG

Coja5452_c0_seq1	*00000016036	Dopey family member 2	0.04	0.34	**1.86**	P, GE	N	0.01	0.09	0.11	0.56	U	FG
Coja15494_c0_seq1	*00000002159	Acyl-Coenzyme A oxidase 1, palmitoyl	0.04	0.25	0.48	P	N	0.00	0.15	0.03	1.00	U	L
Coja9711_c0_seq1	*00000001782	Phospholipase A2-activating protein	0.04	0.25	**0.52**	P, GE	N	0.01	0.14	0.09	0.91	U	
Coja16265_c0_seq1	*00000012543	Uncharacterized protein	0.04	0.23	0.63	P	N	0.01	0.17	0.09	1.00	U	T
Coja14546_c0_seq1	*00000001055	Uncharacterized protein	0.04	0.22	**0.62**	P, GE	N	0.01	0.11	0.08	0.99	U	
Coja16548_c0_seq1	*00000002899	Acetoacetyl-CoA synthetase	0.04	0.18	**0.93**	P, GE	N	0.00	0.14	0.03	1.00	U	L
Coja14863_c0_seq1	*00000006775	Transglutaminase 2 (C polypeptide, protein-glutamine-gamma-glutamyltransferase)	0.04	0.16	-0.44	P	N	0.04	0.30	0.13	0.99	U	
Coja3118_c0_seq1	Not ortholog		0.04	0.15	-0.09	P	N				0.44	U	
Coja12455_c0_seq1	*00000010555	Oxysterol-binding protein	0.04	0.14	**-0.56**	P	N	0.01	0.14	0.09	1.00	U	
Coja11734_c0_seq1	*00000010178	Hypothetical protein LOC42735	0.04	0.14	-0.36	P	N	0.02	0.16	0.12	1.00	U	
Coja14997_c0_seq1	*00000003883	Myotubularin related protein 1	0.04	0.13	0.48	P	N	0.01	0.14	0.08	1.00	U	
Coja16065_c0_seq1	*00000001684	SEC14-like 1 (S. cerevisiae)	0.04	0.12	**0.65**	P, GE	N	0.00	0.11	0.02	0.97	U	T
Coja15471_c0_seq1	*00000009445	Inositol polyphosphate-5-phosphatase F	0.04	0.10	**0.71**	P, GE	N	0.01	0.10	0.09	0.64	U	
Coja8178_c0_seq1	*00000012006	Reticulon 1	0.04	0.05	-0.24	P	N	0.02	0.09	0.20	0.92	M	T
Coja21502_c0_seq1	*00000013123	1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase zeta-1	0.04	0.01	0.34	P	N	0.06	0.13	0.48	1.00	T	T
Coja6460_c0_seq1	*00000011473	Ribosomal protein S6 kinase, 9kDa, polypeptide 2	0.04	0.00		P	N	0.00	0.12	0.01	0.92	T	T

Coja10632_c0_seq1	*00000009328	Eukaryotic translation initiation factor 3, subunit A	0.03	0.65	0.55	P	N	0.00	0.11	0.02	0.57	U	
Coja17497_c1_seq1	*00000006911	Transmembrane channel-like 5	0.03	0.60	**1.75**	P, GE	N	0.06	0.14	0.47	0.99	M	M
Coja5190_c0_seq1	*00000006417	SEC24 family, member A (S. cerevisiae)	0.03	0.55	**0.96**	P, GE	N	0.01	0.12	0.08	1.00	U	
Coja14534_c0_seq1	*00000008057	Lethal giant larvae homolog 2 (Drosophila)	0.03	0.55	**0.67**	P, GE	N	0.02	0.08	0.22	1.00	U	FG
Coja10203_c0_seq1	*00000001965	Alanyl-tRNA synthetase	0.03	0.49	0.43	P	N	0.01	0.18	0.07	1.00	U	
Coja6932_c0_seq1	*00000002478	Matrin 3	0.03	0.47	0.08	P	N	0.01	0.09	0.07	1.00	U	
Coja9203_c0_seq1	Not ortholog		0.03	0.46	-0.28	P	N				0.84	U	
Coja15775_c0_seq1	**00000021305*	*Uncharacterized protein*	0.03	0.41	-0.39	P	Y	0.01	0.13	0.10	1.00	U	
Coja21753_c0_seq1	*00000007273	Uncharacterized protein	0.03	0.37	-0.46	P	N	0.01	0.17	0.07	1.00	U	T
Coja17477_c0_seq1	*00000009252	Phospholipase D1, phosphatidylcholine-specific	0.03	0.30	**1.57**	P, GE	N	0.12	0.45	0.27	0.87	U	FG
Coja25869_c0_seq1	*00000010084	Eukaryotic translation initiation factor 4 gamma, 3	0.03	0.30	0.11	P	N	0.01	0.12	0.10	0.66	U	T
Coja5067_c0_seq1	*00000014869	Nicotinamide nucleotide transhydrogenase	0.03	0.25	0.24	P	N	0.03	0.17	0.16	1.00	U	
Coja17406_c0_seq1	*00000002445	KIAA319-like	0.03	0.24	**0.78**	P, GE	N	0.05	0.17	0.31	0.83	U	T
Coja17361_c0_seq1	*00000006730	Uncharacterized protein	0.03	0.24	**1.32**	P, GE	N				0.93	U	
Coja5970_c0_seq1	*00000008786	Neurobeachin-like 1	0.03	0.21	-0.70	P	N				0.90	U	
Coja15972_c0_seq1	*00000008457	Ras and Rab interactor 2	0.03	0.20	0.03	P	N	0.01	0.11	0.12	0.99	U	
Coja9576_c0_seq1	*00000010061	LPS-responsive vesicle trafficking, beach and anchor containing	0.03	0.17	**1.14**	P, GE	N	0.03	0.14	0.21	0.92	U	FG

Coja11292_c0_seq1	*00000015014	Transportin 1	0.03	0.17	**0.85**	P, GE	N	0.00	0.09	0.00	1.00	U	
Coja5667_c0_seq1	*00000021316	Solute carrier family 12 (potassium/chloride transporters), member 4	0.03	0.13	0.08	P	N	0.00	0.10	0.01	1.00	U	
Coja15032_c0_seq1	*00000006865	Uncharacterized protein	0.03	0.12	0.59	P	N	0.15	0.37	0.42	1.03	U	
Coja12931_c0_seq1	Not ortholog		0.03	0.09	**0.67**	P, GE	N				0.66	U	L
Coja17219_c0_seq1	*00000008953	Aminoadipate-semialdehyde synthase	0.03	0.08	-0.08	P	N	0.01	0.08	0.17	1.00	U	L
Coja11270_c0_seq1	*00000007575	Protein phosphatase 4, regulatory subunit 1-like	0.03	0.06	0.31	P	N	0.02	0.13	0.14	0.98	U	
Coja34242_c0_seq1	*00000006401	Oxysterol-binding protein	0.03	0.05	**-1.07**	P	N	0.05	0.13	0.39	1.03	U	T
Coja8101_c0_seq1	*00000016098	Exocyst complex component 6B	0.03	0.04	-0.11	P	N	0.02	0.15	0.13	1.00	U	T
Coja15422_c0_seq1	*00000014835	Complement component 7	0.03	0.02	**-1.65**	P	Y	0.04	0.13	0.29	1.00	M	L
Coja15596_c0_seq1	Not ortholog		0.03	0.02	-0.30	P	N				0.95	U	T
Coja5860_c0_seq1	*00000008038	Splicing factor 3b, subunit 1, 155kDa	0.02	0.70	-0.08	P	N	0.00	0.13	0.00	1.00	U	U
Coja15873_c0_seq1	*00000015109	Tight junction protein 2 (zona occludens 2)	0.02	0.37	**0.78**	P, GE	N	0.02	0.13	0.11	1.00	U	
Coja16096_c0_seq1	*00000013439	ATPase type 13A1	0.02	0.27	**0.82**	P, GE	N	0.01	0.32	0.03	0.98	U	
Coja17067_c0_seq1	*00000011621	Insulin-like growth factor 2 receptor	0.02	0.23	**1.51**	P, GE	Y	0.03	0.15	0.19	1.00	U	
Coja15984_c0_seq1	*00000008318	STE2-like kinase	0.02	0.20	0.38	P	N	0.02	0.14	0.13	1.01	U	FG
Coja12430_c0_seq1	*00000011995	Cytoplasmic linker associated protein 2	0.02	0.17	0.39	P	N	0.01	0.08	0.09	1.02	U	
Coja17015_c0_seq1	Not ortholog		0.02	0.17	0.02	P	N				0.36	U	U
Coja15053_c0_seq1	*00000003970	Tight junction protein 1 (zona occludens 1)	0.02	0.15	0.20	P	N	0.01	0.10	0.07	0.99	U	

Coja17308_c0_seq1	*00000014388	G patch domain containing 8	0.02	0.09	-0.43	P	N	0.04	0.35	0.11	0.95	U	U
Coja151648_c0_seq1	Not ortholog		0.02	0.08	**1.05**	P, GE	N				0.47	U	
Coja16748_c0_seq1	*00000006829	Uncharacterized protein	0.02	0.08	0.20	P	N				1.00	U	
Coja12498_c0_seq1	*00000016451	Rho-associated, coiled-coil containing protein kinase 2	0.02	0.08	**-0.97**	P	N	0.00	0.08	0.02	0.86	U	
Coja17225_c0_seq1	*00000001565	Complement component 5	0.02	0.00		P	Y	0.04	0.15	0.25	0.99	L	L
Coja10109_c0_seq1	*00000002929	Thyroid hormone receptor interactor 12	0.01	0.34	**0.65**	P, GE	N	0.00	0.07	0.02	1.01	U	T
Coja14876_c0_seq1	*00000016236	Ubiquitin carboxyl-terminal hydrolase	0.01	0.29	0.51	P	N	0.00	0.10	0.00	1.00	U	T
Coja17479_c0_seq1	*00000001546	Zinc finger, ZZ-type with EF-hand domain 1	0.01	0.25	**1.46**	P, GE	N	0.02	0.11	0.21	1.01	U	
Coja17146_c0_seq1	*00000003910	PREDICTED: LOW QUALITY PROTEIN: E3 ubiquitin-protein ligase UBR4	0.01	0.15	-0.23	P	N				0.87	U	
Coja11544_c0_seq1	*00000005660	Neurofibromin 1	0.01	0.08	0.35	P	N	0.00	0.11	0.01	0.79	U	
Coja127586_c0_seq1	*00000002944	Carbamoyl-phosphate synthetase 1, mitochondrial	0.01	0.06	0.01	P	N				0.83	U	
Coja17206_c0_seq1	*00000003170	Dynein, axonemal, heavy chain 1	0.01	0.00	-0.66	P	N	0.05	0.14	0.37	1.01	T	T

[a]Ortholog assigned by reciprocal best BLASTp between quail and chicken protein; annotations are based on chicken ortholog

[b]Entries in italics are based on Ensembl v. 69 annotations that were retired or changed in Ensembl v. 71.

[c]Protein descriptions are based on Ensembl v. 69 gene IDs or Wiki gene IDs

[e]Protein abundance estimated by the exponentially modified protein abundance index (Ishihama, 2005)

[f]Gene expression level estimated as reads per kilobased per million mapped reads

[h]Log fold changes reported from averages of SD vs SD+ T and SD vs LD treatments when acftive; Bold indicates signficant up- or down-regulation

[i]ID Method: P= proteomics, GE = gene expression analyses

[j]Signal Peptide?: Presence of a signal peptide assigned in SignalP 4.1

[k]Pairwise estimates based on quail:chicken orthologs

[l]Ortholog Hit Ratio: Predicted amino acid length of a quail protein divided by the length of the chicken ortholog

[m] En=Enriched: significantly enriched in one tissue by at least log 2 fold change;
Re=Restricted = expressed only in one tissue or ubiquitous; FG = foam gland, L= liver, T = testis, U = ubiquitous

Table S4.4 *P* - values from pairwise Fisher's exact tests comparing the proportion of orthologs among different classes of foam gland expression

Class	GE	P	FG Exp	Other
P + GE	1.1E-14	1.8E-03	3.8E-47	3.3E-98
GE		4.3E-11	8.5E-80	2.2E-291
P			1.2E-74	3.4E-186
FG Exp				4.8E-190

Table S4.5 *P* - values from pairwise Fisher's exact tests comparing the proportion of orthologs across different tissues

Class	Liver	FG	Test	Ubiquitous	P
P+GE	1.0E-107	8.8E-75	1.9E-100	1.2E-34	2.4E-03
Liver		5.1E-01	9.2E-06	5.0E-130	1.7E-167
FG			4.6E-01	1.1E-42	2.8E-92
Test				1.7E-237	4.8E-182
Ubiquitous					1.8E-52

Table S4.6 *P* - values from pairwise Fisher's exact tests performed within each origin class for genes grouped according to relative expression level in foam

Origin	Class	FG Exp	GE	Other	P
Avian					
	GE	2.7E-03	-	-	-
	Other	9.3E-07	5.5E-11	-	
	P	1.6E-01	1.0E+00	3.3E-05	-
	P+GE	1.0E+00	3.0E-01	9.6E-01	3.9E-01
Vertebrate					
	GE	9.6E-01	-	-	-
	Other	2.8E-26	2.0E-19	-	-
	P	1.0E+00	8.1E-01	1.9E-05	-
	P+GE	4.2E-02	2.1E-01	8.5E-09	3.8E-02
Animal					
	GE	6.3E-01	-	-	-
	Other	6.8E-05	1.6E-05	-	-
	P	8.5E-09	1.4E-05	1.4E-14	-
	P+GE	9.4E-02	5.4E-01	8.8E-04	7.2E-01
Eukaryote					
	GE	1.5E-02	-	-	-
	Other	4.1E-23	2.3E-24	-	-
	P	2.1E-01	1.7E-03	1.1E-02	-
	P+GE	1.0E+00	2.5E-01	4.2E-02	1.0E+00
Bac + Arch					
	GE	1.0E+00	-	-	-
	Other	1.5E-07	1.3E-04	-	-
	P	1.8E-13	5.3E-11	5.2E-24	-

Table S4.7 *P* - values from pairwise Fisher's exact tests performed within each origin class for genes grouped according to tissue-specific expression

Origin	Class	FG	Liver	P	P+GE	Test
Avian						
	Liver	1.00E+00				
	P	6.99E-01	5.70E-01			
	P+GE	1.00E+00	1.00E+00	4.87E-01		
	Test	1.00E+00	2.18E-01	8.98E-07	2.78E-01	
	Ubiquitous	1.00E+00	1.00E+00	6.30E-01	1.00E+00	1.15E-12
Vertebrate						
	Liver	1.00E+00				
	P	4.07E-04	5.70E-10			
	P+GE	5.00E-07	2.55E-13	4.78E-02		
	Test	8.17E-02	2.33E-05	9.27E-03	2.07E-06	
	Ubiquitous	3.90E-07	9.55E-21	2.78E-01	4.73E-01	1.50E-16
Animal						
	Liver	1.00E+00				
	P	1.63E-02	7.93E-05			
	P+GE	6.08E-01	1.94E-01	8.11E-01		
	Test	1.00E+00	6.30E-01	2.22E-16	7.18E-05	
	Ubiquitous	1.00E+00	1.00E+00	1.71E-07	2.78E-01	1.05E-08
Eukaryote						
	Liver	3.27E-01				
	P	6.08E-01	1.42E-07			
	P+GE	4.61E-01	1.72E-06	1.00E+00		
	Test	1.00E+00	3.16E-05	2.51E-01	2.84E-01	
	Ubiquitous	5.86E-03	1.10E-17	5.73E-03	6.30E-01	7.30E-16
Bac + Arch						
	Liver	6.99E-01				
	P	3.16E-05	2.08E-04			
	P+GE	1.03E-04	2.10E-03	1.00E+00		
	Test	1.00E+00	3.46E-03	4.62E-28	1.95E-16	
	Ubiquitous	3.52E-01	1.00E+00	1.22E-12	2.64E-06	1.06E-13